

THIS WEEK

EDITORIALS

WORLD VIEW China should end its worship of academics and boost vocational skills **p.381**

MATERIALS A flexible electronic skin that lights up when touched **p.382**

WORTH THE WAIT Tar-pitch droplet captured dripping after 69 years **p.384**

Plastic fantastic

A resurgence in organic technology is set to transform the world of electronic devices, offering a way to give the very fabric of life enhanced functionality.

The transition from basic science to practical technology is rarely linear. The common view that promising discoveries need only patience, hard work and money to attain commercial success is seldom correct. Often, all kinds of technical, economic and social drivers must also coincide. So forecasts of fortune may fail and fade, only for the idea to re-emerge when the climate is more favourable.

Such a resurgence is now under way in organic electronics, in which polymers and other organic molecules are the active materials in information processing. Hideki Shirakawa discovered in the late 1960s that insulating plastics in the form of polyacetylene films could be made to conduct electricity. Chemists Alan Heeger and Alan MacDiarmid collaborated with Shirakawa in 1976 to boost the material's conductivity by doping with halogens, and went on to make a 'polymer battery'.

Greeted enthusiastically by some firms, this early work soon stalled — the polymers were too unstable and difficult to process, and their properties were hard to control and reproduce reliably. The situation changed in the late 1980s, when Richard Friend and co-workers at the University of Cambridge, UK, found that poly(*p*-phenylene vinylene) could conduct without doping and could be stimulated to emit light, paving the way for polymer light-emitting diodes. It began to seem possible that such substances could be used to make lightweight, flexible devices through simple printing and coating techniques.

The synthesis of gossamer-thin organic electronic circuits by Martin Kaltenbrunner at the University of Tokyo and his colleagues (see page 458) is the latest example of the ingenuity driving this field. Their devices elegantly blend new and old materials and techniques. The substrate is a 1-micrometre-thick plastic foil; organic small molecules provide the semiconductor for the transistors; other organic molecules

and alumina make up the insulating layers; and the electrodes are ultrathin aluminium. The plastic films — 27 times lighter than office paper — can be crumpled like paper and stretched to more than double their length without impairing the device's performance. Adding a pressure-sensitive rubber layer produces a touch-sensing foil that could serve as electronic skin or in medical prostheses.

Wearable and flexible devices have recently made great strides, propelled in particular by the work of John Rogers' group at the University of Illinois at Urbana-Champaign. Made from materials that biodegrade safely, such devices can now be printed on or attached directly to human skin. The possibilities for *in situ* monitoring of wound care, tissue repair, brain and heart function, and drug delivery are phenomenal; the challenge is for medical procedures to keep pace with the technology. Such applications show that organic electronics complements silicon logic, taking information processing into areas that silicon will never reach.

These technologies seem potentially transformative — more so on current showing than graphene. The latest work continues the trend towards a smart environment in which all kinds of functionality are invisibly embedded. What happens when clothing, money and even flesh and blood can receive, process and send information; when the fabric of daily life can be turned, unseen, into a computing and sensing device? Most narratives currently dwell on fears of surveillance or benefits of round-the-clock medical checks and diagnoses. But past experience should teach us that technologies don't simply get superimposed on the quotidian — they both shape and are shaped by human behaviour. Whether or not we get what is good for us, it probably won't be what we expect. ■

Antibiotic threat

In the fight to combat antibiotic resistance, researchers should strengthen their advocacy.

The growing threat of antibiotic resistance has seen a notable rise in awareness among policy-makers and the public (see *Nature* **495**, 141; 2013). This is largely because of the advocacy of researchers, who have urged that the problem be tackled immediately. But antibiotic resistance is a multifaceted global issue, and a coordinated international effort will be needed to maintain the pressure to act.

Some relatively simple interventions could impede the march of resistance and buy time for research responses — such as advising physicians not to give unnecessary prescriptions of antibiotics, and

limiting antibiotic use in livestock for growth promotion (see page 398). But experience has shown how difficult it is to change the expectations of patients and physicians, and agriculture lobbies are not going to give up the use of antibiotics without a fight. Attaining even the low-hanging fruit will therefore be a hard-fought battle that will need persistence.

Yet that call is increasingly being heard (see page 394), and the time is ripe for galvanizing sluggish political will. International public- or animal-health agencies and national authorities are now fully aware of the looming threat, and the World Health Organization in particular has been active in sounding the alarm (see go.nature.com/tzwdmz). Furthermore, antibiotic resistance seems to be moving up the agenda of research funders, although much remains to be done to increase the overall level of funding and to ring-fence it — specifically for research into discovering and developing new antibiotics.

Researchers must continue to lobby politicians, funding agencies and the pharmaceutical industry over the need to implement effective means to curtail the rampant spread of resistance, and to address the glaring dearth of new antibiotics in the drug-development pipeline. ■



China needs workers more than academics

As it faces a glut of unemployed graduates despite labour shortages, China should end its worship of qualifications over skills, argues Qiang Wang.

Graduates from Chinese universities are facing the hardest job-hunting season in the nation's history. As of April, seven out of ten students in their final year had not signed an employment contract. Last autumn in the northern Chinese city of Harbin, 2,954 graduates and 29 with a higher degree applied for 457 jobs to be street cleaners.

The situation in China seems typical of a worldwide trend in which the economic slump has seen the title of 'graduate' become almost synonymous with 'soon-to-be unemployed'. Yet this bleak outlook for Chinese college graduates is happening as the country heads towards a period of labour shortage. The legacy of China's one-child policy combined with recent economic growth has created a demand for workers and a dwindling supply.

In Guangdong province, the shortfall of skilled factory workers rose to more than a million people this year. Meanwhile, Chinese teenagers continue to enrol in universities, lured by the promised rewards of an academic qualification in science or other subjects.

The Chinese people have long believed that higher education is a way to reach the top of the country's pyramid-shaped social structure and join the elite ranks of officialdom: a Confucian slogan for education is "he who excels in study can follow an official career". But whereas academic education is valued, vocational education is held in deep and wide contempt. Vocational students are usually seen as the losers of education, even by themselves and their parents. Furthermore, the Chinese government sees US innovation as a product of that country's education system and has set US higher education — with its heavy focus on academic achievement over useful skills — as an example to follow. Yet China's production-oriented economy does not need more graduates with pieces of paper to proclaim how clever they are. It needs more well-trained workers.

China began a centrally planned college-development programme in 1998. The number of undergraduate students soared from just over one million that year to almost seven million in 2013. The target was to develop world-class academic universities that would be judged by their ranking in national and international league tables of research performance. Many Chinese vocational schools, which previously trained students to prepare for blue-collar work, were merged into these kinds of university. In a single generation, China's higher-education system shifted from being measured by the skill of the students it sent into society to being measured by the quantity and quality of the academic papers it produced. This violent tilt towards academic and away from vocational education has contributed to the improvement of Chinese scholarly papers — the number of

articles published by China-based authors in *Nature*-branded journals, for example, increased from 12 in 2000 to 303 in 2012. However, this deliberate unbalancing of higher education has also created a generation of students who are over-educated and under-skilled.

Many young Chinese people with a degree are in jobs that do not require one. And students with master's and PhD degrees are finding it even harder than graduates with lower degrees to find employment. Statistics from the Department of Education of Guangdong show that the employment rate of postgraduates in the region has been lower than that of undergraduates since 2006. At the same time, a shortage of skilled workers in China has led to salary inflation. Between 2009 and 2010, the labour cost in the Pearl River delta and Yangtze River delta, two Chinese manufacturing centres, has risen by 20–25%, and nationally the average minimum wage has risen by 12%. In 2012, migrant

workers earned an average of 2,290 renminbi (US\$374) per month, a higher salary than 69% of that year's college graduates.

If the situation is not addressed, the problem will grow more severe. China should look not to the United States but to Germany, which has a long tradition of balancing academic and vocational education. In Germany, more than half of all high-school students go into technical and vocational education and training, which integrates work- and school-based learning to prepare apprentices for full-time employment. Most German universities focus more on teaching than on research, which tends to take place in independent institutes that are embedded in academic clusters.

Through creating well-trained employees, German higher education is widely recognized as contributing not only to the success of Germany's export-dependent economy, but also to a low rate of youth unemployment (7.6%). By comparison, over the past 5 years in the United States, youth unemployment has risen more than in previous recessions and is currently more than 16%. It should also be pointed out that German manufacturing workers make around 30% more per hour than those in the United States (\$47 compared with \$36).

The revival of vocational education is crucial to rebalance higher education in China. Community colleges should become vocational schools in manufacturing centres, and private investments — guaranteed by the government — should be used to build new vocational schools.

China must no longer look down its nose at vocational education. The Chinese economy needs it, and the country's bright young people deserve more than a certificate and a handshake for their efforts. ■

**CHINA MUST
NO LONGER
LOOK DOWN
ITS NOSE AT
VOCATIONAL
EDUCATION.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/x9pnpf

Qiang Wang is director of the Western Research Center for Energy and Eco-Environmental Policy at the Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China.
e-mail: qiangwang7@gmail.com

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

ELECTRONICS

Skin that glows on touch

A flexible electronic skin, or e-skin, lights up instantly in response to touch.

A team led by Ali Javey at the University of California, Berkeley, manufactured the e-skin by layering carbon nanotube transistors, light-emitting diodes (LEDs) and pressure sensors. The authors made a 3.5- by 3-centimetre patch of e-skin that lights up where pressure is applied — the higher the local pressure, the brighter the LEDs. The researchers suggest that the technology could be modified to respond to sensations other than pressure and may have applications in interactive displays, robotics and health monitoring.

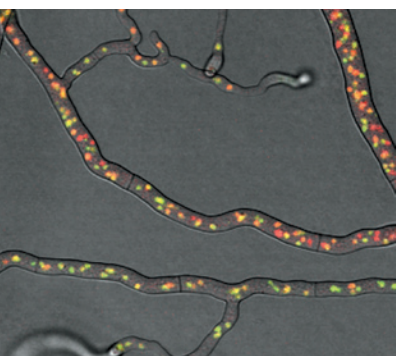
Nature Mater. <http://dx.doi.org/10.1038/nmat3711> (2013)

FUNGAL BIOLOGY

Mould optimized for gene mixing

Some moulds boost their genetic diversity by pumping DNA through sprawling networks of fungal fibres.

Unlike plants and animals, certain fungi form colonies of interconnected cells with mobile, genetically distinct nuclei in a common cytoplasm. Marcus Roper at the University



of California, Los Angeles, and his colleagues tagged nuclei of the red bread mould *Neurospora crassa* with either green or red fluorescent proteins (**pictured**) and then monitored how the nuclei moved through fungal filaments, which branch and fuse into a hyphal network. The team found that fluids push nuclei through the tiny tubes at a rate of 3 millimetres per hour, supplying the tips of the network with a diverse mix of DNA. The network's structure enhances hydraulic flow, boosts nuclear mixing and could help fungi to recombine genes to create more robust mould, the authors say.

Proc. Natl Acad. Sci. USA
<http://dx.doi.org/10.1073/pnas.1220842110> (2013)

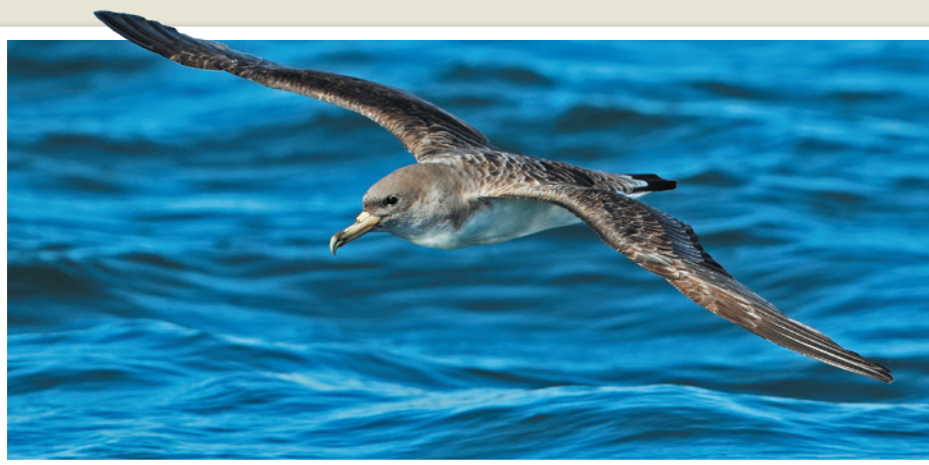
ANIMAL BEHAVIOUR

Sea birds scent home

Scent helps Cory's shearwaters to navigate long distances over featureless oceans.

Anna Gagliardo at the University of Pisa in Italy and her team deprived eight shearwaters (*Calonectris borealis*; **pictured**) of their sense of smell by washing their nasal cavities with a zinc sulphate solution. The authors attached magnets to the heads of eight more birds to disrupt any natural magnetic sense, and used another eight not subject to sensory manipulation as controls.

All animals were tagged and released around 800 kilometres from their home colony. The control birds and seven of those carrying magnets returned to the colony within a few days. Just two of the smell-deprived birds made it home within the breeding period, and only after following long, tortuous paths. The researchers suggest that shearwaters navigate by an odour-based map, and not by geomagnetic fields. **J. Exp. Biol.** 216, 2798–2805 (2013)



ROBIN CHITTENDEN/FLPA

CANCER RESEARCH

Tumour lines are not all equal

The cell lines that researchers routinely turn to when studying ovarian cancer are not the best genetic match available.

Nikolaus Schultz and his team at the Memorial Sloan-Kettering Cancer Center in New York screened 47 cell lines used to model ovarian cancer. The researchers compared the models with samples of human ovarian cancer in terms of gene expression, genetic duplications and mutations. Many frequently used lines were quite different from the most common and lethal cancer types. For example, two cell lines most often used as

tumour models did not contain mutations in a gene that is typically altered in ovarian cancers. Less-used lines were deemed better matches. The analysis will help researchers to choose the most-relevant lines for testing potential cancer drugs, the authors say, and the approach could be used for other tumour types.

Nature Commun. 4, 2126 (2013)

ASTRONOMY

Stellar ice hints at planet birth

A 'snow line' in the gas-rich disk around a young star shows how far from the star carbon monoxide freezes, and so where planets are likely to form. Carbon monoxide ice

PATRICK HICKEY/UCLA

SEVEN DAYS

The news in brief

POLICY

Embryo bans lifted

On 16 July, the French National Assembly approved legislation to permit research on human embryonic stem cells and embryos. The decision, which fulfils a 2012 campaign pledge by President François Hollande, was backed largely by the socialist majority and opposed by conservatives. Such research was previously banned, with some exceptions for work promising “major therapeutic progress” for serious diseases (see *Nature* **469**, 277; 2011). Although the ban has now been lifted, research will be strictly regulated by the French Biomedicine Agency.

Carbon tax scrapped

Australia will shift from a carbon tax to an emissions trading system for greenhouse gases one year ahead of schedule, announced Prime Minister Kevin Rudd on 16 July. The government said that it was abandoning the controversial tax, which took effect last year, to reduce costs for consumers (see go.nature.com/9afdee). The move is expected to lower the price of carbon from about US\$23 a tonne to around \$6 a tonne beginning in July 2014, and should save an average family some \$350 per year.

Regulators saved

UK agencies that regulate human-embryo and human-tissue research have been spared the axe. After initially planning to dissolve the Human Fertilisation and Embryology Authority and the Human Tissue Authority, the government suggested earlier this year that the two bodies could be merged. But it has now accepted the recommendations of an independent review released



Tar drip drops

After 69 years of waiting, scientists in Ireland captured the first video footage of a tar-pitch droplet's plop earlier this month. The experiment, started in 1944 at Trinity College Dublin, demonstrates the fluid nature of tar, or asphalt, which appears to be solid at room temperature. Physicists are eager to study the details of a droplet's escape for the first time. Although an identical experiment has been going on in Queensland, Australia, since 1927, cameras missed its most recent drip in 2000. The next plunge of an Australian drop is expected by the end of 2013. See go.nature.com/wbngvw for more.

on 17 July, which found that there is little overlap between the work of the agencies and that they should remain separate entities. See go.nature.com/f7nt6k for more.

EPA delay ends

The US Senate voted on 18 July, after a 133-day delay, to confirm Gina McCarthy as administrator of the Environmental Protection Agency (EPA). The wait is the

longest faced by any chief in the EPA's 43-year history (see *Nature* **497**, 418–419; 2013). Senate Republicans had used procedural tactics to delay formal votes on McCarthy and on several other nominees proposed by President Barack Obama's administration to head federal agencies. They relented when Democrats threatened to change Senate rules. At the EPA, McCarthy will oversee regulations

to limit greenhouse-gas emissions and to improve water and air quality.

Marine-reserve veto

Proposals to create two huge marine reserves in Antarctic waters were blocked by Russia on 16 July at an international meeting in Bremerhaven, Germany. Plans put before the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR), which oversees fishing activity in Antarctica, would have banned fishing in 1.6 million square kilometres of the Ross Sea and created seven protected areas off the coast of East Antarctica. Despite the failure to reach agreement, CCAMLR members are expected to continue discussing the plans in Hobart, Australia, in October. See go.nature.com/usayrh for more.

EU pesticide ban

The European Union (EU) is set to ban the insecticide fipronil from use in agricultural fields, owing to concerns that the chemical is contributing to the drastic decline in Europe's bee population. On 16 July, 23 member states supported the restriction, with 2 opposing and 3 abstaining. From next year, fipronil use will be mainly limited to seeds sown exclusively in greenhouses. The European Food Safety Authority in Parma, Italy, concluded in May that maize (corn) seeds treated with fipronil pose a “high acute risk” to honeybees.

iPS trial approved

On 19 July, Japan's health minister, Norihisa Tamura, approved the world's first trial in humans of induced pluripotent stem (iPS) cells. Masayo Takahashi, a stem-cell biologist at the RIKEN Center for Developmental

DAVID WHYTE

Biology in Kobe, plans to use sheets of retinal cells derived from iPS cells to repair retinal epithelium in patients with age-related macular degeneration, a common cause of blindness. Final approval by Takahashi's institution is expected soon, and she plans to start recruiting patients as early as September.

EVENTS

China tremors

Two earthquakes rocked an area near the city of Dingxi in Gansu province, northwest China, on 21 July. The tremors — one at magnitude 5.9, followed by one at magnitude 5.6 — struck a region that has seen three quakes greater than magnitude 6 in the past four decades. News services reported at least 89 deaths, along with major structural damage to buildings. The relatively shallow depth of both quakes — about 10 kilometres — may have contributed to the damage.

PEOPLE

Turing pardon

Four years after issuing a formal apology, the UK government has cleared the way to grant a posthumous pardon to the British mathematician Alan Turing (pictured), who in 1952 was convicted of 'gross indecency'



under anti-homosexuality legislation (see *Nature* 482, 441; 2012). He later took his own life. On 19 July, the government offered its support for a bill that would overturn the conviction. Turing is regarded as a hero of the Second World War for helping to break the German Enigma code. His 'Turing machine' concept is considered to underlie modern computer science.

RESEARCH

Cosmic gas guzzler

An immense gas cloud several times the mass of Earth has started its death spiral around the supermassive black hole at the heart of the Milky Way. Observations from the European Southern Observatory's Very Large Telescope in Chile, revealed on 17 July, show the cloud swinging and stretching around the black hole. Extreme gravitational forces

are expected to rip apart the cloud, called G2, in the coming months. See go.nature.com/3kpqzj for more.

India trials halted

The US National Institutes of Health (NIH) in Bethesda, Maryland, last week confirmed reports that it has postponed some of its roughly 30 clinical trials in India, and has stopped enrolling participants in others. In January, the Indian health ministry tightened regulations on clinical trials. Sponsors are now required to provide compensation to participants who are injured as a result of a trial, or to the surviving relatives of those who are killed. "NIH has expressed its concerns about the new regulations, and looks forward to hearing clarifications from the Indian government," the agency told *Nature*.

BUSINESS

Pharma probe

The Chinese government is investigating four senior executives of GlaxoSmithKline (GSK) in China for allegedly bribing officials and physicians to boost GSK drug sales, funnelling the money through travel agencies. The executives are suspected of giving 3 billion renminbi (US\$489 million) in bribes since 2007, Chinese officials said last week. The government

COMING UP

28 JULY–1 AUGUST

Scientists discuss conflicts between humans and wildlife at the 50th annual conference of the Animal Behavior Society in Boulder, Colorado. go.nature.com/qjdgheh

29 JULY–6 AUGUST

The American Physical Society will discuss plans for high-energy physics at a meeting in Minneapolis, Minnesota (see page 391). go.nature.com/7yiv5p

29 JULY–2 AUGUST

The meteor that exploded over Russia in February is on the agenda at the Meteoritical Society's annual meeting in Edmonton, Canada. go.nature.com/gtbpqv

says that the probe is part of nationwide crackdown to ensure fair competition in the pharmaceutical industry. GSK, headquartered in London, says that it is conducting its own review and is cooperating with the Chinese police.

GM crops dropped

Agricultural biotechnology giant Monsanto has abandoned efforts to win regulatory approval for the cultivation of new genetically modified (GM) crops in the European Union (EU). The company confirmed last week that it is withdrawing all pending EU applications for new transgenic maize (corn), soya beans and sugar beet. The firm, based in St Louis, Missouri, said that it would focus instead on its conventional agriculture business in the EU. See page 387 and go.nature.com/vkaxty for more.

➔ NATURE.COM

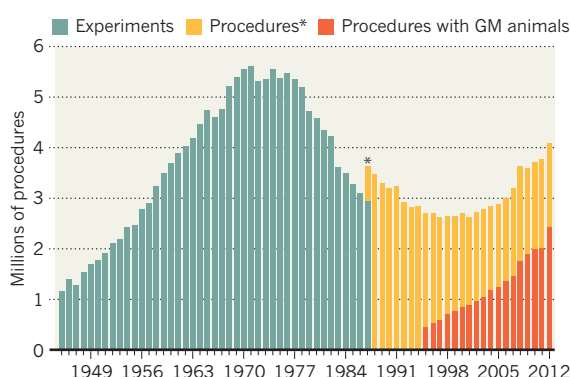
For daily news updates see: www.nature.com/news

TREND WATCH

The number of research procedures involving animals in the United Kingdom has continued to rise, reaching 4.11 million in 2012, according to Home Office data released last week. The 8% increase over 2011 figures was driven largely by greater use of genetically modified and other mutant animals. Mice accounted for 76% of the 4.03 million animals that were used for the first time last year in procedures including breeding and experiments. See go.nature.com/t9z6uk for more.

UK ANIMAL EXPERIMENTS ON THE RISE

Research procedures involving animals are at a 20-year high, fuelled by increasing use of genetically modified (GM) animals.



*UK switched to recording procedures in 1987. Creating a GM animal and conducting an experiment may be considered separate procedures. No GM data before 1995.

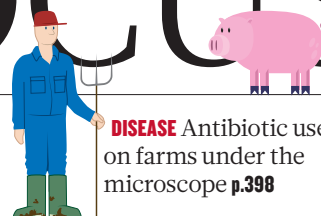
NEWS IN FOCUS

REGENERATIVE MEDICINE Doubt raised over existence of tiny stem cells **p.390**

PARTICLE PHYSICS New life for thrifty, old-school cyclotrons **p.391**

CONSERVATION China wall planned to kill off plant pest **p.392**

DISEASE Antibiotic use on farms under the microscope **p.398**



Hostility towards the cultivation of transgenic crops in Europe has prompted vociferous protests.

BIOTECHNOLOGY

Monsanto drops GM in Europe

Region abandoned owing to stalled approval process.

BY DANIEL CRESSEY

European researchers have expressed regret but little surprise at last week's announcement by the agriculture giant Monsanto that it will no longer be seeking approvals for genetically modified (GM) crops now under review for cultivation in the European Union (EU).

As anti-GM campaigners celebrated, advocates warned that Europe risks becoming a scientific backwater as the rest of the world increasingly adopts the technology. "It's bad news for Europe, for European farmers and for global food security," says Jonathan Jones, who uses both GM and conventional approaches to study disease resistance in plants at the Sainsbury Laboratory in Norwich, UK. "Europe has to get its act together."

The approval process for GM crops has ground to a halt in Europe despite a clear regulatory path. Crops must first be deemed safe by the European Food Safety Authority (EFSA) in Parma, Italy. The European Commission must then produce a draft decision within three months, to be voted on by representatives from EU member states before approval can be finalized.

The EFSA has deemed eight crops as safe, some as long ago as 2005. But political disquiet over the cultivation of GM crops, including bans in some EU countries, has meant that the commission has not moved forward on any of them.

➔ **NATURE.COM**
For more on the promise and reality of GM crops, see: go.nature.com/gafvrp

Four crops in limbo — three varieties of maize (corn) and one

of soya bean — are Monsanto products. The company, which is based in St Louis, Missouri, also has five GM crops still under review by the EFSA: four maize varieties and one sugar-beet variety. Monsanto says that it will abandon applications for all of them except for one GM maize, MON810. This is already grown in the EU, but is now up for its ten-year reapproval review. That reapproval has already been passed to the commission by the EFSA.

Monsanto will now focus its European efforts on its conventional agriculture business and on enabling the import of GM crops for use as animal feed, a widespread EU practice that is less controversial than cultivating the crops in European fields.

MON810, which is modified to produce a protein called Bt that is harmful to insect pests, is one of only two GM crops approved for cultivation in the EU. The other is a high-starch GM potato called Amflora that is intended for industrial applications such as paper production. But in 2012, its makers, BASF Plant Science in Limburgerhof, Germany, announced that, because of hostility to GM products in Europe, it would abandon development and sale of Amflora and move its base to the United States. The company also later said that it will not pursue EU approvals for other GM potatoes.

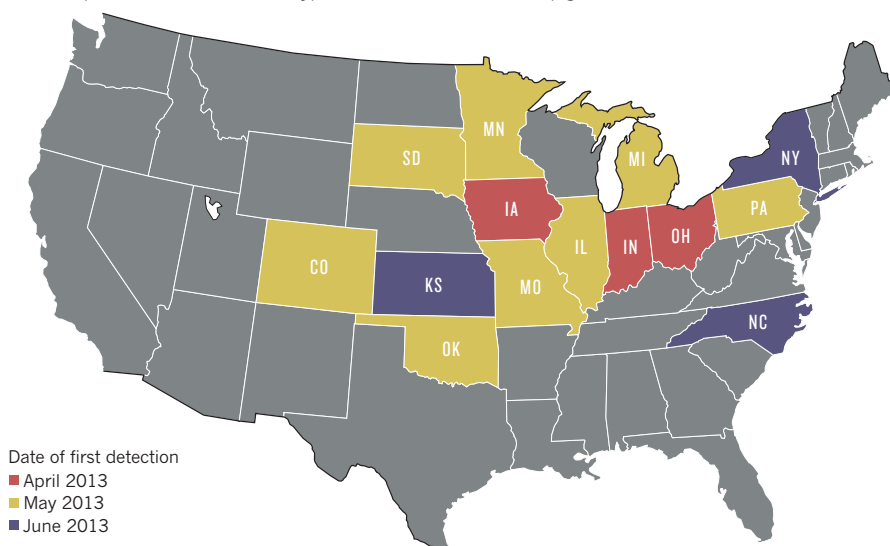
EFSA applications for four GM maize varieties from other companies are still pending — two from DuPont Pioneer, based in Johnston, Iowa, and two from Syngenta in Basel, Switzerland. Pioneer says that it will continue with its approvals, and Syngenta says that its position is also unchanged.

Maurice Moloney, director of Rothamsted Research, an agricultural-research centre in Harpenden, UK, says that the Monsanto move is a "perfectly reasonable business decision", given the current political climate. Although he expects that Europe will eventually come round politically, for now, even small-scale field trials can attract vociferous protests. Increasingly, researchers are looking for the chance to do their work in more accepting countries in the Far East, Africa and Latin America, says Denis Murphy, a plant biotechnologist at the University of South Wales near Pontypridd, UK. "I do a lot of my work now overseas," he says. "I've almost given up on Europe." ■

Additional reporting by Richard Van Noorden.

PIG VIRUS ON THE WING

Porcine epidemic diarrhoea virus, a type of coronavirus that can kill piglets, has been detected in 14 US states.



ANIMAL DISEASE

Deadly pig virus slips through US borders

Researchers race to track spread of coronavirus.

BY BETH MOLE

A lethal virus that causes diarrhoea and vomiting in pigs has entered the United States and has been found in 14 states. With the country's US\$97-billion pork industry standing to lose millions of dollars in the event of a mass outbreak, scientists are working to track the virus and prevent its spread, even as they try to understand how it passed through biosecurity defences in the first place.

"How this virus got here, that's the million-dollar question," says James Collins, director of the Veterinary Diagnostic Laboratory at the University of Minnesota in St Paul.

The pathogen, a type of coronavirus called porcine epidemic diarrhoea virus (PEDV), was first identified in the United Kingdom in 1971, and it caused mass epidemics in Europe in the 1970s and 1980s. As pigs there developed immunity, the virus petered out and now causes only occasional, isolated outbreaks. It has since spread to Asia, where it has been considered endemic since 1982, causing substantial economic losses to pork producers. The virus can spread quickly by a faecal-oral route and infect entire herds. And although adult pigs typically recover, PEDV can kill 80–100% of the piglets it infects. The virus poses no health threat to humans.

The US Department of Agriculture (USDA)

had tried to keep PEDV and other diseases out of the country by restricting imports of pigs and pork products from certain nations, such as China. But on 10 May, the Veterinary Diagnostic Laboratory at Iowa State University in Ames confirmed that PEDV had infected pigs in Iowa, the leading producer of US pork. The lab then screened samples taken earlier from other states and found a case from Ohio submitted on 16 April that is now the earliest known US detection of PEDV, according to Gregory Stevenson, a pathologist at Iowa State. The fact that the virus has now spread to 14 states in total is a sign that the outbreak is still flaring and could become an epidemic (see 'Pig virus on the wing').

"It's a real threat," says Lisa Becton, a veterinary surgeon and director of swine health information at the National Pork Board, an industry group in Des Moines, Iowa.

To understand the virus's enigmatic US entry, scientists are sequencing viral DNA isolated from pigs and comparing it with PEDV variants from elsewhere in the world. Researchers are also trying to create rapid diagnostic tests and vaccines to prevent the virus from spreading. The National Pork Board has approved \$800,000 to fund research and education.

But PEDV must first be grown in labs — a notoriously difficult exercise because the pathogen thrives in the specific conditions found in

pig guts. Researchers in Europe and Asia have already managed to infect cells, but only after years of working with the virus. In the United States, the same import restrictions that were set up to help prevent PEDV from entering the country have made it difficult to import the necessary lab materials for working with the virus, such as vaccines, infected cells and pig antibodies.

"What's hampering the research is that we don't have reagents," says Linda Saif, a virologist at Ohio State University in Wooster. Access to the virus and good tests in hand "would have helped us identify which herds have been exposed, and one could have imposed more stringent control measures", she says.

The USDA's National Veterinary Services Laboratories in Ames is one of just a few US facilities to have grown the virus successfully. But it had a head start: the lab imported the virus around 15 years ago from Asia, after a lengthy security-clearance process, in preparation for just such an outbreak. Lab scientists have spent recent months tweaking cell-culture protocols, and plan to distribute the virus to researchers on request in the coming weeks.

In the meantime, other research groups have focused on detecting viral DNA in sick pigs and on sequencing viral genes. In August, a team led by Douglas Marthaler, a scientist at the University of Minnesota's Veterinary Diagnostic Laboratory, will publish the sequence of a virus genome taken from a Colorado farm. They found it to be 99.4% identical to a Chinese strain of PEDV. On the basis of that sequence, many researchers suspect that the virus originated in China, but Marthaler says that he is surprised by the level of similarity, because he would have expected the US virus to have evolved more in the time since it arrived.

In any case, he says, the potential origin of the virus does not say anything about the route that it took to reach the United States. Canada, the main source of pigs entering the United States, does not import pigs from China either. And although researchers know that the virus can be transported in faeces, they do not know how long it can survive outside pigs' intestines, so it is unclear if a dirty boot, a contaminated package or an illegal import carried PEDV into the country.

Vets say that pig farmers are now restricting access to farms, and are cleaning pig manure more carefully off their clothes and trucks as they move between barns. And researchers still hope that they can elucidate the virus's international and domestic path by looking for subtle evolutionary changes in viral genome sequences of samples from Asia and different US states.

Saif, who has feared such an outbreak for decades, wonders what the virus will do next. Agriculture experts speculate that it may be more stable in cooler temperatures, and thus more dangerous, making the current outbreak a mild precursor to what could come in the winter. "We have to be vigilant," says Saif. ■

SOURCE: US DEPARTMENT OF AGRICULTURE

MEDICINE

Stricter standards sought to curb stem-cell confusion

Initiative aims to clarify description of mesenchymal cells.

BY HELEN SHEN

Pamela Robey is used to being sent samples by scientists who are anxious to know whether the mesenchymal stem cells (MSCs) they have extracted from fat can be coaxed to turn into either bone or cartilage.

Robey, who directs the Stem Cell Unit at the US National Institutes of Health (NIH), is also used to delivering bad news to many of those who seek her help. "They usually are not happy," she says, when her attempts to differentiate the cells produce little more than fatty globules.

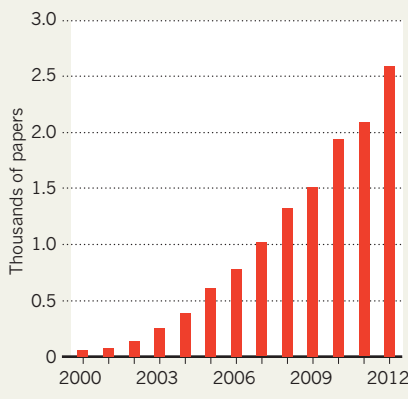
To Robey, that disappointment reflects a pervasive misunderstanding about what MSCs are and what they can do — one that is fuelled by a lack of information. MSCs have been proposed as treatments for a wide range of ailments including heart and brain injury, joint damage, Crohn's disease and multiple sclerosis. But some scientists say that these clinical aspirations have far outrun the basic science. "It always seems a little bit like hocus pocus when you're treating everything from skeletal to immunological disorders," says George Daley, director of the Stem Cell Transplantation Program at Boston Children's Hospital in Massachusetts.

An international group of scientists, industry experts and governmental organizations is trying to introduce scientific clarity to the burgeoning field of MSCs. The group, which met for the first time in late March at NIH headquarters in Bethesda, Maryland, hopes to introduce more rigorous research practices, and eventually to create scientific standards that could guide the commercial development of MSC-based therapeutics. The efforts, applauded by some, are now attracting criticism both from those who advocate MSCs as therapeutics and from those who think that the potential of the cells is oversold.

Even the definition of what constitutes an MSC is a matter of debate. First described in the 1960s as bone-marrow cells that can regenerate bone, the name MSC has expanded to include cells from fat, and from dental and other tissues — although early claims that MSCs could give rise to a panoply of different cell types have fizzled out in animal tests. As a first step towards standardization, the group is developing guidelines for research journals to help them cope with the sharp rise in

GROWTH MEDIUM

The number of papers that mention mesenchymal stem cells has risen dramatically in the past decade.



the number of MSC studies being published (see 'Growth medium'). Ideas being floated include requiring authors to define clearly the animal and tissue sources of their cells, and the experimental conditions used to culture them — details that are often omitted from published papers.

"People have to be much more rigorous in defining MSCs, their sources, which tissues you can obtain them from, and what you can use them for," says Paolo Bianco, a stem-cell researcher at the Sapienza University of Rome, who is not involved in the initiative.

Robey, who is on the steering committee of the new working group, is more blunt. "Most of the MSC biology is not rigorous," she says. "Other stem-cell biologists tend to look down their noses at the field."

CLEAR DEFINITION

The group is also discussing whether authors should quantify what percentage of MSCs in a given preparation truly are stem cells. Tests, including detecting the presence of certain cell-surface molecules, that were once used as baseline criteria for defining MSCs are now known to identify a wide variety of cells, only a tiny fraction of which may be able to self-renew and differentiate into multiple cell types.

The team has not yet approached journals to gauge interest in adopting the guideline plan, and "whether or not they'll accept it, I don't know," says steering-group chairman Armand

Keating, who directs the Cell Therapy Program at Princess Margaret Hospital in Toronto, Canada.

In the coming year, the working group faces the more daunting task of establishing one or more reference lines of MSCs, against which researchers and companies can readily assess the identity, purity and potency of their cells.

But some researchers say that attempts to standardize cells will do little to benefit science, because a solid theoretical basis for using MSCs as broad therapeutics is lacking. "If I want to treat a disease of the brain with cells that for all intents and purposes are bone cells, the first thing I should come up with is a clear rationale," says Bianco.

The group's efforts also worry Arnold Caplan, a stem-cell researcher at Case Western Reserve University in Cleveland, Ohio, who introduced the modern concept of MSCs in 1991 (*A. I. Caplan J. Orthop. Res.* **9**, 641–650; 1991). He believes that the cells could potentially treat a broad range of conditions by secreting biochemicals that reduce inflammation and promote healing, and says that overly prescriptive journal guidelines could prevent the publication of promising results.

It is also hard to know how adoption of more rigorous experimental standards would affect the already strong consumer demand for treatments derived from MSCs. Lee Buckler, founder of the consulting firm Cell Therapy Group, based in Vancouver, Canada, says that introducing standard cell lines and tests into MSC research could make it easier to compare different therapeutic products. "Some companies are going to feel delighted," he says, but adds that more-established businesses may find their products suddenly threatened by competing claims that are founded on clearer points of reference.

If the standards, once arrived at, facilitate the systematic design and evaluation of preclinical trials of MSCs and their products, they could help biotechnology companies to attract large pharmaceutical firms to bring their MSC products to market, says Robert Deans, executive vice-president of Athersys in Cleveland, Ohio, which is developing MSC-derived treatments. Deans, who is participating in the standards working group, says that, until now, "pharma hasn't invested a lot. There's not enough consensus on these cells." ■

Doubt cast over tiny stem cells

Studies refute the existence of very small embryonic-like cells endorsed by the Vatican.

BY ALISON ABBOTT

Does a rare and minuscule cell type with the potential to repair almost any tissue in the body really exist?

Proponents of very small embryonic-like cells (VSELs) extracted from bone marrow say that the cells have the potential to transform regenerative medicine. A trial has begun in Poland, and cell-therapy company Neostem in New York is planning another in Michigan.

But in a major blow to the field, a paper published on 24 July in *Stem Cell Reports* suggests that the diminutive stem cells are not real¹. Led by Irving Weissman, a prominent stem-cell researcher at Stanford University in California, the study is the fourth to refute the cells' existence — and the most thorough yet.

"Weissman's evidence is a clincher — it is the end of the road for VSELs," believes Rüdiger Alt, head of research at Vita 34, a private bank for umbilical cord blood in Leipzig, Germany, who last year published the first failure to replicate claims for the cells².

Robin Smith, chief executive at Neostem, disagrees. She compares the attacks on VSELs to those suffered by Charles Darwin and Nicolaus Copernicus when they proposed their world-changing scientific theories.

The battle over VSELs has been raging for more than two years, and has covered ground from the United States to Vatican City and Poland. The cells were first described³ in mouse bone marrow in 2006, by a team led by Mariusz Ratajczak at the University of Louisville in Kentucky. His group and a few others have since generated a literature that characterizes VSELs as rare components of bone marrow and other tissues, less than 6 micrometres in diameter and able to turn into a diverse range of cell types, including blood, bone, muscle and nerve.

Ratajczak was given a joint position at the Pomeranian Medical University in Szczecin, Poland, in 2006. From there, he obtained €10.6 million (US\$14 million) from European Union sources for a VSEL research network involving five institutions. The network last year registered the first human trial of a VSEL preparation, which aims to treat 60 people who have severe angina. Around one-quarter of the participants have already been injected with the preparation.

But the network became rattled after one collaborator, Józef Dulak at the Jagiellonian University in Krakow, failed to find traces of VSELs

► NATURE.COM
To read about a controversial Italian stem-cell trial, see:
go.nature.com/b3ufcm



Mariusz Ratajczak first reported finding very small embryonic-like stem cells in 2006.

in his experiments. When he published his findings in May⁴, Ratajczak tried to force him out of the consortium. Dulak, like Weissman, found no molecular signatures associated with pluripotency in any mouse bone-marrow cells smaller than 7 micrometres across. Weissman's more extensive analysis now also reports that in his experiments, the small cells did not aggregate into spheres *in vitro*, as pluripotent cells do; nor could they differentiate into blood cells, the adult tissues that such cells are most likely to become.

Alt, for his part, spent 18 months studying human VSELs at Leipzig University. He says that his team gave up when its chromosome analyses revealed that the objects they observed in their experiments were not proper cells. "If they had existed, Vita 34 would have been interested to try to commercialize them," he says. "We wasted a lot of time." A fourth study, which sought VSELs in blood from umbilical cords, could not show that they were pluripotent either⁵.

Together with the University of Louisville, Ratajczak patented his VSEL discovery. In 2007, Neostem acquired exclusive licensing rights and created its Stem For Life Foundation, which promotes adult stem cells as an ethical alternative to embryonic stem-cell therapy. Two years ago, the Vatican, which supports this approach, donated US\$1 million to the foundation — the first time it had entered a joint venture with a trading company. It has since hosted two international conferences on adult stem cells, co-organized

by Stem For Life and Neostem.

Weissman presented some of his now-published results at an unrelated meeting of the Pontifical Academy of Sciences in Vatican City in April 2012. But he says he is annoyed that representatives of the Catholic Church "act as if they don't know the scientific data" and have continued to back VSELs.

"I don't see the controversy — we have seen bone grow" from VSELs in mice, says Russell Taichman, a researcher in dental medicine at the University of Michigan in Ann Arbor. Taichman will be running the Neostem-backed VSEL trial, which will look for bone regrowth in dental patients. Announced in April, the trial is awaiting approval by the US Food and Drug Administration.

The leader of the Polish VSEL trial is also unfazed. "Other investigators are just not managing to catch the right cells," says Wojciech Wojakowski, a cardiologist at the Medical University of Silesia in Katowice.

Diane Krause, who has published evidence that VSELs from mouse bone marrow can differentiate into lung epithelial cells⁶, agrees. "I can only say that we manage to see these cells," she says. "One of our postdocs went to the Ratajczak lab and learnt the technique properly."

Ratajczak thinks that all the published criticisms come down to investigators lacking the technical skill to harvest the correct cells. "Weissman has never visited my lab to witness exactly how we carry out the method," he says. Weissman counters that "science has to be reproducible, and methods described so that others can reproduce them".

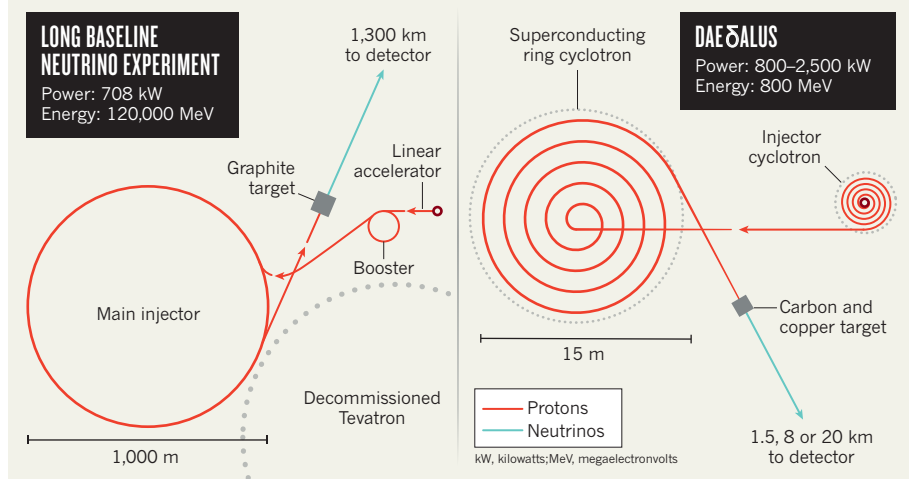
Neostem itself is becoming more cautious in its statements about VSELs. In an e-mail, Smith told *Nature* that the company "has studies in progress to determine, with robust data, whether or not VSELs have characteristics of pluripotent cells".

The debate exemplifies problems often seen in adult stem-cell research, in which a trial can end up being based on inadequate data, says Rudolf Jaenisch, a stem-cell researcher at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts. "It's even worse when it gets the blessing of the church." ■

1. Miyaniishi, M. *et al.* *Stem Cell Rep.* <http://dx.doi.org/10.1016/j.stemcr.2013.01.001> (2013).
2. Danova-Alt, R., Heider, A., Egger, D., Cross, M. & Alt, R. *PLoS ONE* **7**, e34899 (2012).
3. Kucia, M. *et al.* *Leukemia* **20**, 857–869 (2006).
4. Szade, K. *et al.* *PLoS ONE* **8**, e63329 (2013).
5. Alvarez-Gonzalez, C. *et al.* *PLoS ONE* **8**, e67968 (2013).
6. Kassmer, S. H. *et al.* *Stem Cells* <http://dx.doi.org/10.1002/stem.1413> (2013).

CYCLOTRONS RECYCLED

By bringing back cyclotrons, an 80-year-old technology, DAE δ ALUS could challenge the Long Baseline Neutrino Experiment's attempt to look for fundamental matter–antimatter asymmetries using neutrinos.



PARTICLE PHYSICS

Cyclotrons come full circle

US physicists hope to rejuvenate a classic technology to support key neutrino experiments.

BY EUGENIE SAMUEL REICH

Since 1982, particle physicists have sojourned every so often in the fresh mountain air of Snowmass, Colorado, to dream up successive generations of particle accelerators.

Budgetary restrictions mean that this year's Snowmass Community Summer Study — the first since 2001 — will be a bit less ambitious. The grass-roots planning exercise, which begins on 29 July, will last one week rather than three. And physicists will be trading the posh mountain resort for a humbler spot: Minneapolis, Minnesota.

The idea capturing physicists' attention in the run-up to the meeting is correspondingly modest and thrifty: a low-energy but high-intensity accelerator called a cyclotron. The technology, first used more than 80 years ago, eventually ceded ground to the snazzier synchrotrons that today power facilities such as the Large Hadron Collider (LHC) at Europe's particle-physics laboratory, CERN, near Geneva, Switzerland.

But, as physicists look for cheaper ways to test fundamental questions, cyclotrons could experience a renaissance.

"We need something totally out of the

box," says Janet Conrad, a particle physicist at the Massachusetts Institute of Technology in Cambridge, and co-spokesperson for the DAE δ ALUS collaboration, a proposal to generate beams of subatomic neutrinos using linked cyclotrons.

After the US Department of Energy (DOE) in 2011 closed the highest-energy US particle collider, the Tevatron, at Fermi National Accelerator Laboratory (Fermilab) in Batavia, Illinois, US particle physicists focused on creating particle beams that were high in intensity, rather than energy, so that they could generate large amounts of data per second. That way, even if Fermilab would no longer compete with the LHC in terms of energy, it

could repurpose the Tevatron's luminous proton beams and establish an 'intensity frontier'.

Fermilab quickly developed plans for a flagship intensity experiment, called the Long Baseline Neutrino Experiment (LBNE). It would send beams of neutrinos and antineutrinos along a 1,300-kilometre underground path to a detector in the Homestake mine near Lead, South Dakota. Neutrinos and antineutrinos are thought to behave the same in nearly every way, but the LBNE would look for crucial differences. It would monitor the way in which the three different types of neutrino morph, or oscillate, into one another as they travel along the beam, and then repeat the experiment, this time with a beam of antineutrinos. Any detected difference would indicate a fundamental asymmetry — helping to explain why the Universe contains so much more matter than antimatter.

In 2012, the DOE asked Fermilab to strip down these ambitious plans (see *Nature* **485**, 16; 2012). Fermilab came back with a design that would cost about \$800 million in its first phase. But many physicists are concerned that the scaled-back experiment will be less precise.

DAE δ ALUS is now emerging as a lower-cost alternative to test for matter–antimatter asymmetry. In this plan, near-stationary protons would be dumped in the centre of a small cyclotron and accelerated by magnetic fields in spirals until they reached the cyclotron's outer edge. These souped-up protons would then be injected into a second, larger cyclotron, 15 metres across, which would accelerate them further still. The resulting proton beam would then be fired at a carbon and copper target to generate particles called pions, some of which decay into antineutrinos.

Costing about US\$130 million for the cyclotron pair, the DAE δ ALUS scheme would be much cheaper and smaller than the LBNE (see 'Cyclotrons recycled'). The cyclotrons could not match the proton energies of Fermilab's proton accelerators, but, partly by operating at a higher power, they would generate a comparable number of antineutrinos per second, and so produce a similar amount of data. They would also generate a cleaner beam of antineutrinos than the LBNE could, because the energy levels at which DAE δ ALUS operates allow the copper in the target to neatly separate out the positively

MORE ONLINE

MORE NEWS

- Giant viruses hint at 'fourth domain' of life go.nature.com/2m1ab9
- Stem cells reprogrammed using chemicals and no genes go.nature.com/psphw7
- Collections of wild relatives of crops contain large gaps go.nature.com/w4bkob

TOP STORY



Italy publishes its first in-depth evaluation of universities and research centres go.nature.com/zb2gml

SILVIA MORARA/CORBIS

► charged pions that in turn generate the desired antineutrinos.

Pairs of cyclotrons could be set up at three different distances, or baselines, around the same detector, which would measure how many antineutrinos from each beam morph into a different type during flight. By operating simultaneously with three baselines, DAEδALUS could map the shape, or waveform, of the antineutrino oscillations. If there is asymmetry between matter and antimatter, the shape of these oscillations would take on a characteristic pattern — and there would be no need for the experiment to run separately with a beam of neutrinos.

The two experiments would both need large liquid detectors, but Conrad notes that Japan is already planning to build a water-based detector for astrophysical neutrinos that could work with DAEδALUS. The DAEδALUS accelerators are cheap and small enough that they could be reproduced and built up around pre-existing detectors, rather than the other way around. “The beauty of DAEδALUS is that you can forget about accelerators being stuck at laboratories,” says Conrad.

Other physicists are impressed. “It’s an imaginative proposal for how to address neutrino oscillations,” says Jonathan Rosner, a particle physicist at the University of Chicago in Illinois who oversees the Snowmass Study as chair of the American Physical Society’s Division of Particles and Fields.

LBNE co-spokesperson Milind Diwan of Brookhaven National Laboratory in Upton, New York, also says that the idea is technically interesting. But he says it is not yet proven or accepted, and that the United States needs to commit to a strong domestic programme based around the DOE-approved LBNE in order to draw international funding.

If DAEδALUS features prominently in the Snowmass final report, expected in November, it could find a spot in the agenda set by the Particle Physics Project Prioritization Panel (P5), a group of DOE advisers who will convene shortly after receiving the Snowmass report and begin ranking projects. P5 is likely to recommend some US involvement in high-energy machines, both the LHC and a next-generation machine that Japan hopes to host, the International Linear Collider. The panel could also renew an emphasis, set in its 2008 report, on the intensity frontier, which led to the flagship LBNE proposal.

But some US physicists say that shrinking US budgets mean that projects may end up competing more on scientific merit and cost, rather than what they do for the status of any nation’s programme. “We have to tighten our belts,” says Rosner. “It’s not just business as usual.” ■



Cordgrass was originally introduced to China to help reclaim land from the sea, but has become invasive.

CONSERVATION

China’s cordgrass plan is ‘overkill’

Experts say Shanghai sea wall will cripple wetland habitat.

BY JANE QIU IN SHANGHAI

On an island in Shanghai at the mouth of the Yangtze River in China, a modern-day Great Wall is planned to thwart invaders. This time, however, the enemy is a marauding weed — and the US\$163-million wall will not keep the invaders out, but enclose and drown them.

The plant in question is *Spartina alterniflora*, or cordgrass, a fiercely invasive salt-marsh grass that is native to eastern North America. In 1979, the plant was introduced to China because its roots can trap sediment, making it ideal for erosion control and land reclamation. Since then, the weed has spread rapidly across the country’s coastal regions, covering some 400,000 hectares (see ‘Coastal invasion’). Cordgrass has infested many countries, but “the world’s largest invasion is in China,” says Donald Strong, an ecologist at the University of California, Davis. “It’s a piece of ecological engineering that has gone out of control.”

The threat of cordgrass is especially acute on Chongming Island, home on its eastern end to the 24,000-hectare Shanghai Chongming Dongtan National Nature Reserve — declared a region of international ecological importance by the Ramsar Convention, a global wetland-conservation treaty. Millions of migratory birds

overwinter here, and it is a precious spawning and feeding ground for more than 60 fish species — including the critically endangered Chinese sturgeon (*Acipenser sinensis*).

Cordgrass has been spreading there “like a cancer,” says reserve director Tang Chendong, so far consuming more than 10% of the wetland. It colonizes an area by forming dense bundles with deeply penetrating roots — squeezing out native plants and burrowing invertebrates, eating mudflats and drying up wetlands. At Dongtan, “this has had devastating consequences for many bird species,” says Ma Zhijun, an ornithologist at Fudan University in Shanghai. The reed parrotbill (*Paradoxornis heudei*) — which nests on native reeds and is one of Chongming’s best-known species — was once abundant in the reserve, for example. But its numbers have declined by 40% in the past decade, with more than half of its habitat now overtaken by cordgrass. The shrinking of mudflats also threatens the hundreds of thousands of migratory shorebirds that rely on the reserve as a stopover site.

The best way to get rid of cordgrass over a large area is to spray herbicides from a helicopter — as has been successfully done in the western United States, Australia and New Zealand. But the reserve’s management could not get permission to fly helicopters over

► charged pions that in turn generate the desired antineutrinos.

Pairs of cyclotrons could be set up at three different distances, or baselines, around the same detector, which would measure how many antineutrinos from each beam morph into a different type during flight. By operating simultaneously with three baselines, DAEδALUS could map the shape, or waveform, of the antineutrino oscillations. If there is asymmetry between matter and antimatter, the shape of these oscillations would take on a characteristic pattern — and there would be no need for the experiment to run separately with a beam of neutrinos.

The two experiments would both need large liquid detectors, but Conrad notes that Japan is already planning to build a water-based detector for astrophysical neutrinos that could work with DAEδALUS. The DAEδALUS accelerators are cheap and small enough that they could be reproduced and built up around pre-existing detectors, rather than the other way around. “The beauty of DAEδALUS is that you can forget about accelerators being stuck at laboratories,” says Conrad.

Other physicists are impressed. “It’s an imaginative proposal for how to address neutrino oscillations,” says Jonathan Rosner, a particle physicist at the University of Chicago in Illinois who oversees the Snowmass Study as chair of the American Physical Society’s Division of Particles and Fields.

LBNE co-spokesperson Milind Diwan of Brookhaven National Laboratory in Upton, New York, also says that the idea is technically interesting. But he says it is not yet proven or accepted, and that the United States needs to commit to a strong domestic programme based around the DOE-approved LBNE in order to draw international funding.

If DAEδALUS features prominently in the Snowmass final report, expected in November, it could find a spot in the agenda set by the Particle Physics Project Prioritization Panel (P5), a group of DOE advisers who will convene shortly after receiving the Snowmass report and begin ranking projects. P5 is likely to recommend some US involvement in high-energy machines, both the LHC and a next-generation machine that Japan hopes to host, the International Linear Collider. The panel could also renew an emphasis, set in its 2008 report, on the intensity frontier, which led to the flagship LBNE proposal.

But some US physicists say that shrinking US budgets mean that projects may end up competing more on scientific merit and cost, rather than what they do for the status of any nation’s programme. “We have to tighten our belts,” says Rosner. “It’s not just business as usual.” ■



Cordgrass was originally introduced to China to help reclaim land from the sea, but has become invasive.

CONSERVATION

China’s cordgrass plan is ‘overkill’

Experts say Shanghai sea wall will cripple wetland habitat.

BY JANE QIU IN SHANGHAI

On an island in Shanghai at the mouth of the Yangtze River in China, a modern-day Great Wall is planned to thwart invaders. This time, however, the enemy is a marauding weed — and the US\$163-million wall will not keep the invaders out, but enclose and drown them.

The plant in question is *Spartina alterniflora*, or cordgrass, a fiercely invasive salt-marsh grass that is native to eastern North America. In 1979, the plant was introduced to China because its roots can trap sediment, making it ideal for erosion control and land reclamation. Since then, the weed has spread rapidly across the country’s coastal regions, covering some 400,000 hectares (see ‘Coastal invasion’). Cordgrass has infested many countries, but “the world’s largest invasion is in China,” says Donald Strong, an ecologist at the University of California, Davis. “It’s a piece of ecological engineering that has gone out of control.”

The threat of cordgrass is especially acute on Chongming Island, home on its eastern end to the 24,000-hectare Shanghai Chongming Dongtan National Nature Reserve — declared a region of international ecological importance by the Ramsar Convention, a global wetland-conservation treaty. Millions of migratory birds

overwinter here, and it is a precious spawning and feeding ground for more than 60 fish species — including the critically endangered Chinese sturgeon (*Acipenser sinensis*).

Cordgrass has been spreading there “like a cancer,” says reserve director Tang Chendong, so far consuming more than 10% of the wetland. It colonizes an area by forming dense bundles with deeply penetrating roots — squeezing out native plants and burrowing invertebrates, eating mudflats and drying up wetlands. At Dongtan, “this has had devastating consequences for many bird species,” says Ma Zhijun, an ornithologist at Fudan University in Shanghai. The reed parrotbill (*Paradoxornis heudei*) — which nests on native reeds and is one of Chongming’s best-known species — was once abundant in the reserve, for example. But its numbers have declined by 40% in the past decade, with more than half of its habitat now overtaken by cordgrass. The shrinking of mudflats also threatens the hundreds of thousands of migratory shorebirds that rely on the reserve as a stopover site.

The best way to get rid of cordgrass over a large area is to spray herbicides from a helicopter — as has been successfully done in the western United States, Australia and New Zealand. But the reserve’s management could not get permission to fly helicopters over

Chongming, says Tang, and authorities in Shanghai feared that herbicides would damage native plants, wildlife and local fisheries. So Chinese researchers have come up with another approach. After years of studying cordgrass in field experiments, ecologist Li Bo at Fudan University and his colleagues found that if the plants are cut to the ground and submerged for six months, nearly all of them die.

That is where the wall comes in. In October, construction of an 8-metre-high, 27-kilometre-long sea wall is set to begin at the heart of the reserve. It will stop tides reaching a 2,400-hectare, cordgrass-ridden area that will be subdivided into self-contained grids by mud dykes, with each grid being treated separately. Since early July, researchers have been cutting down cordgrass to test how grid size, topography and tidal patterns might affect eradication efficiency.

Walls are not new to Chongming Island, which sits a few dozen kilometres from the teeming city of Shanghai. Every inch of land is precious, and sea walls have been used in past decades to reclaim land from the sea for agriculture and other human uses.

Conservation experts say that the cordgrass-eradication scheme — given US\$163 million by the Shanghai Forestry Bureau and approved in 2008 — is intriguing, but some think that it is overkill. Financially, “it’s an incredible, major project” to get rid of some weeds, says David Melville, a conservationist in Nelson, New Zealand, who is advising the reserve about the project. The cordgrass “is certainly a problem”, he says. “It’s just a matter of whether one really needs to build such a massive, permanent sea wall to resolve it.”

Melville and other experts note that an

8-metre-high wall — accounting for more than 80% of the project’s budget — is unnecessary when cordgrass can be killed with just half a metre of water. Crucially, they worry that a permanent barricade will cripple efforts to restore tidal habitats for birds. For example, without the ebb and flow of tides, the sea bulrush (*Scirpus mariqueter*), a native intertidal grass bearing fruit and stalks that are key food sources for many birds, will not be able to grow back, says Wang Tianhou, a wetlands ecologist at East China Normal University in Shanghai.

“It’s unrealistic to conserve habitats without any compromise.”

More fundamentally, the current plan will not even eradicate all cordgrass from the reserve, says Li. Following approval in 2008, the project was

put on hold until after the World Expo 2010 in Shanghai because officials did not want to compromise the city’s image with a building project inside a nature reserve. During that time, the cordgrass grew beyond the proposed sea-wall borders by a few hundred metres.

Cordgrass “will continue to eat up the tidal flats and spread to other parts of the reserve”, Li says. But a scheme with a different wall and a larger enclosure would cost more and require another round of approval that could take years, he says — so he feels it is best that the project goes ahead.

Some researchers argue that temporary mud dykes could be a solution: they would cost much less than a sea wall, be easier to secure approval for, and could eradicate the cordgrass in the entire reserve while retaining natural tidal rhythms. But “factors other than conservation were likely to be involved in the decision-making process” that led to the plan for a permanent sea wall, says Ma. As more farmland is converted to hotels, amusement parks and golf courses in Chongming, local officials are increasingly pressed to find lands to replace them — as required by the central government. Reclaiming land from the sea is seen as an easy and cheap option.

Researchers such as Li are pragmatic about the possibility that part of the reserve might one day be co-opted for farming. “It’s unrealistic to conserve habitats without any compromise,” he says. “The officials need to see that they will get something out of the project, too.” In the meantime, the scheme will at least ameliorate the cordgrass infestation — and will create more habitats for birds. “It’s a better option if the alternative is to do nothing,” Li says. ■

CORRECTION

The News story ‘Rinderpest research restarts’ (*Nature* **499**, 264–265; 2013) wrongly referred to the Institute of Animal Health. This has now changed its name to the Pirbright Institute.

SOURCE: LI BO



THE LAST RESORT

Health officials are watching in horror as bacteria become resistant to powerful carbapenem antibiotics — one of the last drugs on the shelf.

BY MARYN MCKENNA

As a rule, high-ranking public-health officials try to avoid apocalyptic descriptors. So it was worrying to hear Thomas Frieden and Sally Davies warn of a coming health “nightmare” and a “catastrophic threat” within a few days of each other in March.

The agency heads were talking about the soaring increase in a little-known class of antibiotic-resistant bacteria: carbapenem-resistant Enterobacteriaceae (CREs). Davies, the United Kingdom’s chief medical officer, described CREs as a risk as serious as terrorism (see *Nature* **495**, 141; 2013). “We have a very serious problem, and we need to sound an alarm,” said Frieden, director of the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia.

Their dire phrasing was warranted. CREs cause bladder, lung and blood infections that can spiral into life-threatening septic shock. They evade the action of almost all antibiotics — including the carbapenems, which are considered drugs of last resort — and they kill up to half of all patients who contract them. In the United States, these bacteria have been found in 4% of all hospitals and 18% of those that offer long-term critical care. And an analysis carried out in the United Kingdom predicts that if antibiotics become ineffective, everyday operations such as hip replacements could end in death for as many as one in six¹.

The language used by Davies and Frieden was intended to break through the indifference with which the public usually greets news about antibiotic resistance. To close observers, however, it also had a tinge of exasperation. CREs were first identified almost 15 years ago, but did not become a public-health priority until recently, and medics may not have appreciated the threat that they posed. Looking back, say observers, there are lessons for researchers and health-care workers in how to protect patients, as well as those hospitals where CREs have not yet emerged.

“It is not too late to intervene and prevent these from becoming more common,” says Alexander Kallen, a medical epidemiologist at the CDC. At the same time, he acknowledges that in many places, CREs are here for good.

Hindsight is key to the story of CREs, because it was hindsight that identified them in the first place. In 2000, researchers at the CDC were grinding through analyses for a surveillance programme known as Intensive Care Antimicrobial Resistance Epidemiology (ICARE), which had been running for six years to monitor intensive-care units for unusual resistance factors. In the programme’s backlog of biological samples, scientists identified one from the Enterobacteriaceae family, a group of gut-dwelling bacteria. This particular sample — of *Klebsiella pneumoniae*, a common cause of infection in intensive-care units — had been taken from a patient at a hospital in North Carolina in 1996 (ref. 2). It was weakly resistant to carbapenems, powerful broad-spectrum antibiotics developed in the 1980s.

Antibiotics have been falling to resistance for almost as long as people have been using them; Alexander Fleming, who discovered penicillin, warned about the possibility when he accepted his Nobel prize in 1945. Knowing this, doctors have used the most effective drugs sparingly: careful rationing of the powerful antibiotic vancomycin, for example, meant that bacteria took three decades to develop resistance to it. Prudent use, researchers thought, would keep the remaining last-resort drugs such as the carbapenems effective for decades.

The North Carolinian strain of *Klebsiella* turned that idea on its head. It produced an enzyme, dubbed KPC (for *Klebsiella pneumoniae* carbapenemase), that broke down carbapenems. What’s more, the gene that encoded the enzyme sat on a plasmid, a piece of DNA that can move easily from one bacterium to another. Carbapenem resistance had arrived.

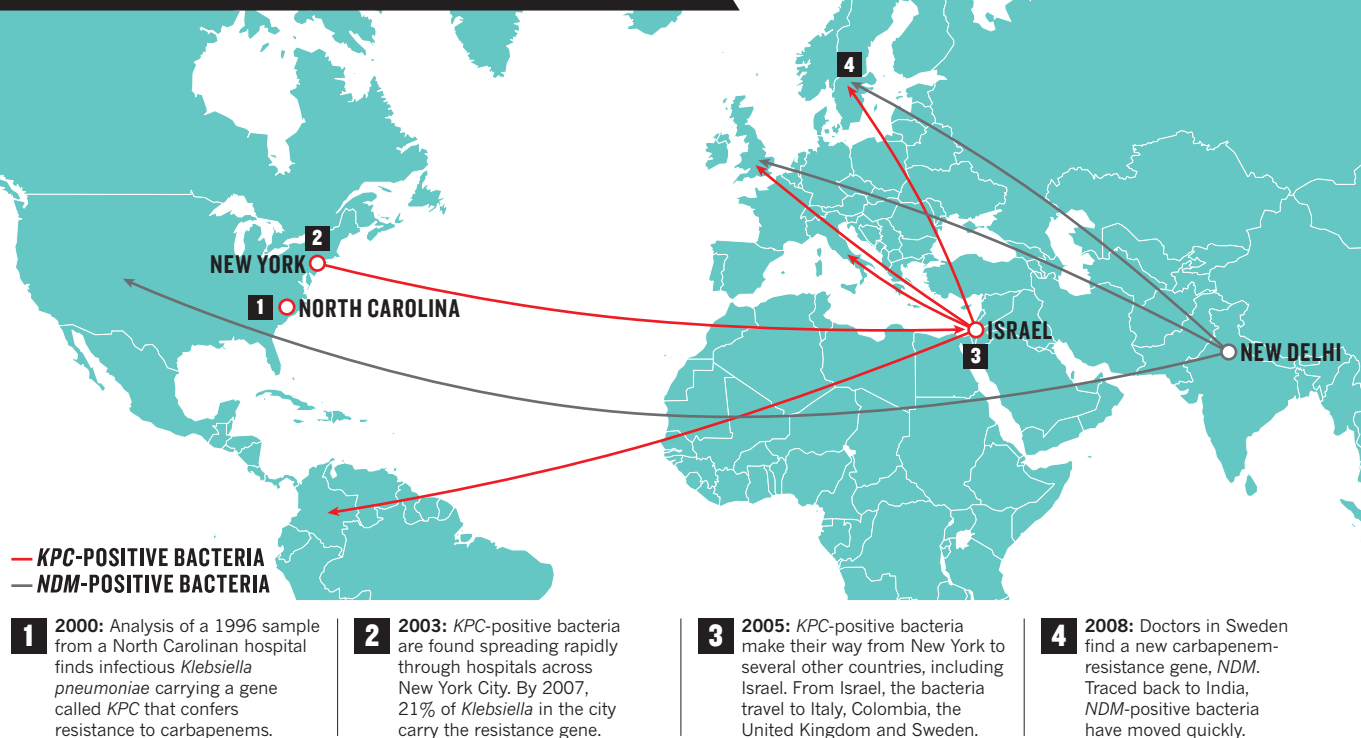
At first, however, microbiologists considered this CRE to be a lone case. Jean Patel, a microbiologist who is now deputy director of the CDC’s office of antimicrobial resistance, says that CDC staff were reassured by the fact that the sample had been collected four years earlier and that testing of the remaining archives revealed no further instances of resistance. “It wasn’t that there was a lack in interest in looking for these,” Patel says.

➔ **NATURE.COM**

For a podcast about carbapenem resistance, see: go.nature.com/dxmtfny

THE RESISTANCE MOVEMENT

Carbapenem-resistant Enterobacteriaceae have been on the move since at least 1996.



Instead, the attitude at the time was, “We have a system for identifying these and it’s working, and if more occur we’ll hear about it”.

But the CDC’s surveillance programme was limited: it tracked only 41 hospitals out of some 6,000 and its analyses lagged far behind sample collection. So when carbapenem resistance emerged again, years passed before anyone noticed.

A DIRE TREND

The State University of New York (SUNY) Downstate Medical Center in Brooklyn draws patients from some of the poorest neighbourhoods in New York City, so it tends to be a place where dire health trends surface. It was not part of the CDC’s ICARE programme, but physicians there conduct their own bacterial surveillance to scan for emerging infectious threats. In 2003, a review of results from the centre’s microbiology lab and some collaborating ones at nearby hospitals picked up something that city physicians had never seen before. Over the previous six years, a handful of patients spread across seven institutions had been diagnosed with *Klebsiella* infections that were partially resistant to carbapenems. “These had been infrequent and they were flying under the radar,” says John Quale, a medical researcher at Downstate. “And at about the time we picked them up, they just exploded.”

The infections were very serious. In one Brooklyn hospital outbreak, 9 out of 19 patients died. In another, two infections blossomed into more than 30 in just six months, despite stringent infection-control measures. And the organism spread around the city — from Harlem Hospital at the north end of Manhattan to Mount Sinai Hospital on the Upper East Side, and then to Saint Vincent’s in Greenwich Village in the south, where one patient died of a *Klebsiella* infection despite doctors throwing every drug they could at it.

One of the reasons why the resistant strains spread so rapidly was that they were difficult to detect. Most clinical microbiology labs no longer painstakingly culture bacteria over days to determine which drugs they are susceptible to: instead, automated systems, which expose bacteria to graduated dilutions of drugs, can give a result in hours. But these tests,

Quale and his collaborators realized, were giving misleading results and were causing physicians to give patients doses or drugs that would not work. And because the infections were not eliminated, the resistant strain could be passed on. By 2007, 21% of all *Klebsiella* bacteria in New York City carried the carbapenem-resistance plasmid, compared with an average of 5% across the rest of the United States³.

Such a rapid dissemination hinted that CREs were travelling from person to person rather than arising independently in each location. This made sense. Many Enterobacteriaceae, including *Klebsiella*, reside in the intestines and can easily be carried by an asymptomatic patient. If patients develop diarrhoea, as often happens after the administration of drugs during intensive care, the infectious bacteria can spread far, contaminating equipment or the hands of care-givers inside the hospital and out. So it was easy to imagine how CREs might ride the subway from Brooklyn to Manhattan. But it took a few years, and a much larger outbreak, to illustrate just how far CREs had travelled (see “The resistance movement”).

RAPID SPREAD

In late 2005, one patient at Tel Aviv’s Sourasky Medical Center was diagnosed with a KPC-positive infection that was closely related to a New York strain. Within months, CRE infections stormed through the hospital, and then through Israel’s small, tight-knit health-care system. By March 2007, there were 1,275 cases nationwide⁴. They were turning up across a network of hospitals, nursing homes, dialysis clinics and rehab centres.

Israel has a shortage of acute-care beds, explains Mitchell Schwaber, an infection-control physician who was on the Sourasky faculty when the KPC epidemic began. “Whenever a patient can be discharged, especially from internal medicine, they are — which creates a lot of movement from acute-care facilities to long-term care facilities, and then back to either the same hospital or a different one.”

In response, the Israeli Ministry of Health created a national task force on CREs, headed by Schwaber. It demanded daily national-surveillance reports by e-mail, and instituted strict isolation precautions, including

dedicated wards, equipment and nurses. The new rules were backed up by surprise inspections and mandatory lab analyses to ascertain where new infections were coming from.

By mid-2008, Israel had reversed its soaring trend of resistant *Klebsiella*. But control came too late to prevent the pathogen from emigrating: patients, physicians and nurses had brought bacteria carrying the KPC enzyme to Italy, Colombia, the United Kingdom and beyond.

ALARM CALL

In January 2008, a urine culture performed on a sample from a 59-year-old man hospitalized in Sweden identified a *K. pneumoniae* strain that was resistant to multiple drugs, including carbapenems⁵. But rather than using KPC, the bacterium dismantled the antibiotics with a different enzyme, a metallo- β -lactamase. Within three years, more cases involving bacteria carrying this enzyme were identified in the United Kingdom and in the United States. These provoked immediate alarm: they were even more resistant to carbapenems than the KPC-carrying *Klebsiella* bacteria, and included other Enterobacteriaceae such as *Escherichia coli*.

Initially, most individuals carrying bacteria with the new resistance factor had some link to clinics in India, through medical tourism or health care needed while abroad.

In accordance with taxonomic conventions, doctors named the new enzyme New Delhi metallo- β -lactamase (NDM), after the place where the initial Swedish patient was thought to have picked it up. The name proved unexpectedly controversial: Indian media and the Indian parliament denounced the acronym for stigmatizing India's medical-tourism industry. Further work by the team that first identified NDM only increased the outrage when it established that bacteria carrying the enzyme were present in sewage and municipal water in south Asia⁶ (see *Nature* <http://doi.org/dgcs33>; 2011).

The controversy obscured NDM's real significance: not only had another resistance mechanism emerged, but CREs were now flourishing beyond hospital walls.

Researchers were still struggling to pin down exactly how NDM was spreading. In the second half of 2012, staff at the University of Colorado Hospital in Aurora discovered that their institution had unknowingly hosted eight patients with NDM-positive *Klebsiella* bacteria, the largest cluster in the United States so far. The first three cases, all in patients with pneumonia, were found during a routine review of clinical specimens. When the hospital escalated its search, it also identified five asymptomatic carriers.

"There was no obvious pattern," recalls Michelle Barron, the hospital's infection-control physician. "These patients had been in the hospital a long time. They had been on multiple units. There was no single piece of equipment that had been used on all of them."

Even when the CDC sequenced the bacterial genomes from all eight patients, the data could not explain how the bacterium had spread. Barron hypothesizes that, at some point, the hospital harboured a "ghost patient" — someone who escaped detection despite the surveillance dragnet. She is still looking for that person: the hospital is attempting to call back and sample all 1,700 patients who were treated during the crucial period.

The episode ended well. The five carriers never fell ill, the three who were ill recovered, and once the hospital became aware of the cluster no further spread occurred. They might not be so lucky next time.

Fresh threats are on the way. Researchers have spotted other carbapenem-resistance factors moving around the globe; one has already appeared in the United States, and others are clustered in southern Europe and South America. Because each is genetically different, they are likely to present new challenges to detection. Infectious-disease specialists say that they have learned major lessons from CREs. Drug-resistant bacteria can emerge and spread much faster than patchy

public-health surveillance systems and outdated laboratory-detection methods can pick them up, and what seems like adequate infection control cannot always contain their spread.

Some countries are trying to take those lessons on board. Hospitals in Israel now practise 'active surveillance', meaning that if a new patient has been to any other health-care institution in the past six months they are checked for CREs. And anyone who tests positive for such bacteria is flagged as a carrier in national-health records, which are accessible to hospitals, nursing homes and community physicians. France and the United Kingdom follow similar rules, but unfortunately many countries do not. Earlier this month, the European Centre for Disease Prevention and Control in Stockholm published a candid self-assessment by 39 European countries of their CRE burden and ability to counter these organisms⁷. Only 21 said they have achieved the kind of national coordination that allowed Israel to contain its epidemic.

The United States operates a patchwork of surveillance systems. The CDC looks for CREs through three separate data networks, but none of these covers the entire country. At least nine states have made reporting CRE cases to their health departments mandatory. The CDC has also created a robust tool kit of best practices for health departments

and hospitals, such as restricting staff assignments and equipment use in hospitals, and identifying infections in the long-term care facilities that feed patients into hospitals. These measures helped institutions in Illinois and Florida to shut down outbreaks in 2008 and 2009.

LIMITED OPTIONS

Meanwhile, lab-detection methods have improved; the CDC's use of whole-genome sequencing to solve the Colorado episode was the first time that the agency deployed that technology to tackle a hospital outbreak. And public-health departments' ability to identify threats has been bolstered by boluses of federal money after the 2001 World Trade Center attack and subsequent anthrax attacks, and in the 2009 stimulus package. But these investments might be rolled back during the current federal budget sequester.

Physicians who treat patients unlucky enough to be caught up in these outbreaks have no better medicines than they did when CREs first emerged. Some organisms respond to two drugs, tigecycline and colistin (also known as polymyxin E). Neither works in every patient, and colistin is notorious for damaging the kidneys. Physicians find themselves caught between using bad drugs or using no drugs at all.

It seems unlikely that new drugs will become available soon. Conversely, the rapid advance of resistance and the consequent need to use these drugs sparingly has convinced pharmaceutical companies that antibiotics are not worth the investment.

That means, say infectious-disease experts, that their best tools for defending patients remain those that depend on the performance of health personnel: handwashing, the use of gloves and gowns, and aggressive environmental cleaning. Yet even research that could improve best practices has been short-changed, says Eli Perencevich, an infectious-diseases physician and epidemiologist at the University of Iowa in Iowa City who studies how resistant bacteria move around hospitals. "We haven't invested in research in how to optimize even standard infection-control practices. We just blame the health-care workers when they go wrong." ■ [SEE EDITORIAL P.379](#)

Maryn McKenna is an independent journalist in Atlanta, Georgia.

1. Smith, R. & Coast, J. *Br. Med. J.* **346**, f1493 (2013).
2. Yigit, H. et al. *Antimicrob. Agents Chemother.* **45**, 1151–1161 (2001).
3. Hidron, A. I. et al. *Infect. Control Hosp. Epidemiol.* **29**, 996–1011 (2008).
4. Schwaber, M. J. et al. *Clin. Infect. Dis.* **52**, 848–855 (2011).
5. Yong, D. et al. *Antimicrob. Agents Chemother.* **53**, 5046–5054 (2009).
6. Kumarasamy K. K. et al. *Lancet Infect. Dis.* **10**, 597–602 (2010).
7. Glasner, C. et al. *Eurosurveillance* **18**, 28, art. 3 (2013).



Pigs carrying methicillin-resistant *Staphylococcus aureus* were found on US farms for the first time in 2007.

FARMING UP TROUBLE

Microbiologists are trying to work out whether use of antibiotics on farms is fuelling the human epidemic of drug-resistant bacteria.

BY BETH MOLE

The sight of just one boot coming through the doorway cues the clatter of tiny hoofs as 500 piglets scramble away from Mike Male. “That’s the sound of healthy pigs,” shouts Male, a veterinarian who has been working on pig farms for more than 30 years. On a hot June afternoon, he walks down the central aisle of a nursery in eastern Iowa, scoops up a piglet and dangles her by her hind legs. A newborn piglet’s navel is an easy entry point for bacterial infections, he explains. If this pig were infected, she would have an abscess, a lump of inflamed tissue, just below the navel. “In human terms, she’d be an outie instead of an innie,” he says, rubbing the pig’s healthy, pink belly button.

Nearly six years ago, an outbreak of ‘outies’ at this nursery marked the first known infection with methicillin-resistant *Staphylococcus aureus* (MRSA) in pigs in the United States. MRSA has troubled hospitals around the world for more than four decades and has been infecting people outside of health-care settings since at least 1995

(see *Nature* **482**, 23–25; 2012). It causes around 94,000 infections and 18,000 deaths annually in the United States. In the European Union, more than 150,000 people are estimated to contract MRSA each year. Its first appearance on a US farm signalled the expansion of what many believe is a dangerous source of human infection.

Male investigated the infections with Tara Smith, an epidemiologist at the University of Iowa in Iowa City, who has since launched one of the most comprehensive investigations yet of where MRSA lives and how it spreads into and out of agricultural settings. She has surveyed farms and grocery stores as well as people’s homes, noses and pets. Her findings could help to end a raging debate about whether farms’ use of antibiotics is contributing to the rise of drug-resistant bacterial infections in humans.

Scientists and health experts fear that it is, and that drug-resistant bacteria from farms are escaping via farmworkers or meat. Last year, the US Food and Drug Administration (FDA) recommended more restraint

DANIEL ACKER/BLOOMBERG/GETTY

in the use of antibiotics in livestock, following the lead of regulatory authorities in other countries (see *Nature* **481**, 125; 2012).

But the meat and agricultural industries are fighting those restrictions. They claim that MRSA and other drug-resistant bacteria that cause human infections arise in hospitals, and that meat production includes safety measures, such as sanitation rules in slaughterhouses, that prevent resistant bacteria from spreading to and infecting people. “There’s a long way between the farm and the table,” says Ron Phillips, a representative for the Animal Health Institute, a trade organization based in Washington DC that represents veterinary-medicine companies.

The major problem has been lack of data. Many farmers are reluctant to allow scientists access to their facilities, and farmworkers — many of whom, in the United States, are undocumented immigrants — are wary of anyone who might want to sample them. But Smith and a small group of researchers are starting to fill the void. They have “really shaped the state of knowledge in the United States”, says Christopher Heaney, an epidemiologist at Johns Hopkins University in Baltimore, Maryland. Smith’s current research, says Heaney, could allow officials to “truly say where these bacteria in people’s noses are coming from”.

PROFIT AND LOSS

At a concentrated animal-feeding operation (CAFO) about an hour’s drive west from Ames, Iowa, the usual din of the nursery is punctuated by the sound of piglets sneezing, thanks to an outbreak of H1N2 influenza. Craig Rowles, a veterinarian and the farm’s manager, surveys his charges, some of which have mucus dripping from their snouts. “It’s just like when you bring kids to a day-care centre,” he says. “After a while, they’re going to come home with a snotty nose.”

Rowles is using a vaccine to fight the outbreak, but he is also dosing the pigs with two antibiotics — chlorotetracycline and Denagard — to prevent secondary bacterial infections. The combination is also routinely used to prevent bacterial diarrhoea and other common ailments in piglets.

Such practices have been common for decades. But few CAFOs have veterinarians on staff to advise on antibiotic use. They are not required to: veterinary antibiotics are generally available over the counter. And some CAFO operators use antibiotics much more liberally than Rowles does. Small doses of antibiotics in feed curb low-grade infections that might otherwise stymie livestock growth. Studies have found that certain antibiotics can increase pigs’ growth rate by 2.5%, enough to make the difference for farmers between profit and loss. In the current US market, a farmer might get around US\$1 per pound for a pig that costs about \$0.94 per pound to produce. Although farm owners do not always reveal the quantities or types of antibiotics they use, an analysis of FDA data by researchers at the Johns Hopkins Center for a Livable Future in Baltimore found that in 2009, some 13.1 million kilograms — 80% of the antibiotics sold in the United States that year — were used on farms.

Antibiotic use on such a broad scale leads to resistant microbes. In a 1976 study, Stuart Levy, a microbiologist at Tufts University School of Medicine in Boston, Massachusetts, found that when farmers started using tetracycline, the numbers of tetracycline-resistant bacteria on the farms spiked¹. Within months, resistance had spread to microbes in farmworkers’ intestinal tracts. “You don’t have to look that far to see resistant bacteria moving to the environment,” Levy says.

In humans, *S. aureus* generally lives peacefully on the skin and in the nose. But if the bacterium enters the body through a wound, for example, it can become an aggressive pathogen and eventually make its way into the bloodstream to cause deadly infections. Most infections succumb to antibiotics, but resistant varieties, including MRSA, can be difficult if not impossible to cure. There are 270,000 strains, each potentially harmful.

Smith and her colleagues are distinguishing the strains of *S. aureus* around Iowa City in part by sequence type (ST), a categorization based on DNA sequences from several places in the genome. The sequence type that Male and Smith found in the Iowa nursery in 2007 was ST398 (ref. 2). Before then, researchers had seen ST398 mostly in Europe, where it was found in livestock and farmworkers but usually did not cause infection.

But two years earlier, ST398 had been reported in a hospital in Hong Kong³, from patient samples dating back to the early 2000s. For Smith and her collaborator Lance Price, an epidemiologist at George Washington University in Washington DC, it was a sign that the boundaries between animal and human infections were blurring.

In February 2012, Price, Smith and their colleagues published a genetic analysis of strains related to ST398 isolated from animals and humans around the world⁴. They found that the lineage that gave rise to ST398 originated in humans. At some point, it crossed into livestock, where it acquired genes conferring drug resistance and a preference for pigs before jumping back to humans. As of 2012, ST398 was the cause of up to 20% of human cases of MRSA in the Netherlands, although the infections are generally mild.

“No one has even looked at these strains in the United States,” says Smith. Doctors, she says, often don’t determine what strain of MRSA is causing an infection, so it is possible that the bug has been stealthily migrating between farms and hospitals for years.

Smith’s next step was to see if there were other ways MRSA might make it off the farm. At a grocery store on the outskirts of Iowa City, Smith pulls a shopping cart from a metal corral, takes a sterile swab from her purse, wipes the cart’s handle and deposits the swab in a plastic sheath. Then she heads to the meat case. “I think the average consumer doesn’t think of this as a risk,” Smith says, picking up a shrink-wrapped tray of bright red steaks. Her swabs have told a different story.

Beginning in January 2012, Smith and her research assistant, Dipendra Thapaliya, spent a year collecting weekly swabs and meat samples from local grocery stores, including this one. They found *S. aureus* on nearly every type of surface. Five per cent of grocery carts carried MRSA. Of meat samples, 30% harboured *S. aureus*, 11% had *S. aureus* resistant to multiple antibiotics and 3% carried MRSA. The data, which have not been published, also showed that pork

products had some of the highest levels of MRSA, whereas meat labelled ‘antibiotic free’ had little or none. This mirrors what Smith and her colleagues found in samples from farms across the state.

ON THE MENU

Now Smith is conducting detailed genetic analyses of the samples to identify MRSA subtypes and where they might have originated. She and her colleagues have found ST398 in the grocery-store samples. But to their surprise, they also found ST5, which is generally found in hospitals and in ‘community’ infections with no obvious link to farms or hospitals. Last year, Smith and Timothy Frana, a microbiologist and veterinarian at Iowa State University in Ames, found that veterinary students who carried MRSA in their noses after visiting pig farms picked up mainly ST5, most of which was resistant to tetracycline⁵. The presence of ST5 among livestock suggests that *S. aureus* strains may move easily between pigs and people, and may become resistant on farms. “It’s the most interesting finding from our study,” says Frana.

In lab tests, Smith and her colleagues have found that 30% of the *S. aureus* harboured in meat is resistant to tetracycline. Given Levy’s data from the 1970s, this is not surprising, says Smith. But researchers, including Heaney and his group at Johns Hopkins, are finding that a bacterium’s drug-resistance profile can give information about where the bug came from that sequence type and other gene-based

**“IN 2009,
80% OF THE
ANTIBIOTICS
SOLD IN THE
UNITED STATES
WERE USED ON
FARMS.”**

FROM THE FARM AND BACK AGAIN

The liberal use of antibiotics on pig farms has been implicated in the transmission of methicillin-resistant *Staphylococcus aureus* (MRSA), but tracing the origin of an infection is tricky.

Pigs fed antibiotics are known to spread MRSA to farmworkers, and MRSA strains show up in meat products, but it is unclear whether this is a significant source of human infection.

MRSA strains historically associated with hospital and community infections have also turned up on farms.

MRSA infections in humans have been acquired mainly in hospitals or in community settings.

With so much movement, the distinctions between MRSA strains from hospitals, communities and livestock are breaking down.

categorizations may not. In a study comparing workers from different farms, tetracycline-resistant MRSA showed up only in workers from farms where antibiotics were used⁶.

In the last phase of their research, Smith and her colleagues will try to determine whether MRSA is trafficking between farms, households and clinics (see 'From the farm and back again'). They are taking nasal samples from 1,300 people and swabbing doorknobs, kitchen sinks and even family pets in 96 households around Iowa City. The researchers are comparing the *S. aureus* strains they find in these samples with the MRSA strains they have found in grocery stores and farms and with strains local doctors find in infected patients. If a strain shows up in all the locations, the researchers will sequence the whole genomes of individual isolates to retrace their movement and evolution. The results have the potential to create the first complete link between farms and clinical cases of MRSA.

BAN VERSUS BEAST

CAFO supporters acknowledge that farm strains of drug-resistant bacteria could theoretically spread to people. But "I don't see this equating to human health risk," says Scott Hurd, a veterinarian and epidemiologist at Iowa State University who has conducted multiple studies to assess the risk of drug-resistant bacteria spreading through meat production. He says that the average person has a greater chance of dying from a bee sting than of contracting MRSA from pork. Hurd argues that limiting the use of antibiotics on farms could be harmful to human health. Even Smith's grocery-store study found that meat sold as 'antibiotic free' had the highest levels of garden-variety *S. aureus*, suggesting that untreated animals harbour more pathogens. "Animals really do need to be treated," Hurd says.

Nevertheless, regulatory authorities have clamped down on antibiotic use on farms. The European Union began phasing out antibiotics for growth promotion in the late 1990s. Denmark led the charge with a full ban in 2000. (China, however, which claims half the world's pig population, has yet to rein in antibiotic use.)

The bans' effects on drug resistance and human and animal health have been murky. Levy and other supporters of the bans say that the result in Denmark has been positive, pointing to data showing a drop

in the use of antibiotics on farms and an increase in meat production. But opponents, including the Animal Health Institute, point out that the use of antibiotics to treat acute illness in Denmark has increased, as have animal deaths.

Last year, amid mounting pressure from several groups, including the National Resources Defense Council, based in New York, the FDA released new guidance calling for the "judicious use" of antibiotics on farms. The agency discouraged the use of antibiotics for growth promotion and urged label changes to the drugs and more veterinary oversight for their application. Not all the guidelines are yet approved, and compliance is voluntary. Nevertheless, the agency has suggested that it will enforce tougher rules if farmers and drug-makers do not adopt the guidelines within about three years. Few are satisfied with the FDA's policy. Pig farmers and meat-industry representatives consider the move a blow to farmers and animal welfare, and supporters of antibiotic restriction say that the voluntary guidelines do not go far enough. Scientists, meanwhile, have pressed the FDA to reveal more data on how farmers are using antibiotics, so far without success.

Smith, who is concerned that farmers are still overusing antibiotics, hopes that the results of her current research will sway their opinions. Antibiotics on farms can trigger the emergence of resistant strains, she says, and those strains turn up on meat, in grocery stores and in homes, and they can infect people. "For me, that's enough," she says.

At the same time, Smith says that she sympathizes with CAFO operators who are trying to produce meat as safely and efficiently as possible. And although human health should take priority over farm animals, she says, farmers will be reluctant to change until researchers can come up with safe and cost-effective practices to replace the use of antibiotics. For now, Smith says, "we're kind of stumped". ■ [SEE EDITORIAL P.379](#)

Beth Mole is an intern for Nature in Washington DC.

1. Levy, S. B., FitzGerald, G. B. & Maccone, A. B. N. *Engl. J. Med.* **295**, 583–588 (1976).
2. Smith, T. C. *et al.* *PLOS ONE* **4**, e4258 (2009).
3. Ip, M. *et al.* *J. Clin. Microbiol.* **43**, 5069–5073 (2005).
4. Price, L. B. *et al.* *mBio* **3**, e00305-11 (2012).
5. Frana, T. S. *et al.* *PLOS ONE* **8**, e53738 (2013).
6. Rinksy, J. L. *et al.* *PLOS ONE* **8**, e67641 (2013).

COMMENT

EVOLUTION The man who linked ecological isolation and speciation **p.404**



SPACE Mountain guardian asks astronomers to show respect **p.407**

INNOVATION Do Nobel laureates eat more chocolate than the foot soldiers of science? **p.409**

OBITUARY Jerome Karle, crystal structure mathematician, remembered **p.410**

B&C ALEXANDER/ARCTICPHOTO



Pipes transport oil from rigs on Endicott Island in Alaska.

Vast costs of Arctic change

Methane released by melting permafrost will have global impacts that must be better modelled, say **Gail Whiteman, Chris Hope and Peter Wadhams**.

Unlike the loss of sea ice, the vulnerability of polar bears and the rising human population, the economic impacts of a warming Arctic are being ignored.

Most economic discussion so far assumes that opening up the region will be beneficial. The Arctic is thought to be home to 30% of the world's undiscovered gas and 13% of its undiscovered oil, and new polar shipping routes would increase regional trade^{1,2}. The insurance market Lloyd's of London estimates that investment in the Arctic could

reach US\$100 billion within ten years³.

The costliness of environmental damage from development is recognized by some, such as Lloyd's³ and the French oil giant Total, and the dangers of Arctic oil spills are the subject of a current panel investigation by the US National Research Council. What is missing from the equation is a worldwide perspective on Arctic change. Economic modelling of the resulting impacts on the world's climate, in particular, has been scant.

We calculate that the costs of a melting Arctic will be huge, because the region is

pivotal to the functioning of Earth systems such as oceans and the climate. The release of methane from thawing permafrost beneath the East Siberian Sea, off northern Russia, alone comes with an average global price tag of \$60 trillion in the absence of mitigating action — a figure comparable to the size of the world economy in 2012 (about \$70 trillion). The total cost of Arctic change will be much higher.

Much of the cost will be borne by developing countries, which will face extreme weather, poorer health and lower ►



JOSH HANER/THE NEW YORK TIMES/REDUX/EVEVINE

Bubbles of methane emerge from sediments below a frozen Alaskan lake.

► agricultural production as Arctic warming affects climate. All nations will be affected, not just those in the far north, and all should be concerned about changes occurring in this region. More modelling is needed to understand which regions and parts of the world economy will be most vulnerable.

ECONOMIC TIME BOMB

As the amount of Arctic sea ice declines at an unprecedented rate^{4,5}, the thawing of offshore permafrost releases methane. A 50-gigatonne (Gt) reservoir of methane, stored in the form of hydrates, exists on the East Siberian Arctic Shelf. It is likely to be emitted as the seabed warms, either steadily over 50 years or suddenly⁶. Higher methane concentrations in the atmosphere will accelerate global warming and hasten local changes in the Arctic, speeding up sea-ice retreat, reducing the reflection of solar energy and accelerating the melting of the Greenland ice sheet. The ramifications will be felt far from the poles.

To quantify the effects of Arctic methane release on the global economy, we used PAGE09. This integrated assessment model calculates the impacts of climate change and the costs of mitigation and adaptation measures. An earlier version of the PAGE model was used in the UK government's 2006 Stern Review on the Economics of Climate Change to evaluate the effect of extra greenhouse-gas emissions on sea level, temperature, flood risks, health and extreme weather while taking account of uncertainty⁷. The model assesses how the net present value of climate effects

varies with each tonne of carbon dioxide emitted or saved.

We ran the PAGE09 model 10,000 times to calculate confidence intervals and to assess the range of risks arising from climate change until the year 2200, taking into account sea-level changes, economic and non-economic sectors and discontinuities such as the melting of the Greenland and West Antarctic ice sheets (see Supplementary Information; go.nature.com/rueid5). We superposed a decade-long pulse of 50 Gt of methane, released into the atmosphere between 2015 and 2025, on two standard emissions scenarios. First was 'business as usual':

"There is a steep global price tag attached to physical changes in the Arctic."

increasing emissions of CO₂ and other greenhouse gases with no mitigation action (the scenario used by the Intergovernmental Panel on Climate Change Special Report on Emissions Scenarios A1B). Second was a 'low-emissions' case, in which there is a 50% chance of keeping the rise in global mean temperatures below 2°C (the 2016r5low scenario from the UK Met Office). We also explored the impacts of later, longer-lasting or smaller pulses of methane.

In all of these cases there is a steep global price tag attached to physical changes in the Arctic, notwithstanding the short-term economic gains for Arctic nations and some industries.

The methane pulse will bring forward by

15–35 years the average date at which the global mean temperature rise exceeds 2°C above pre-industrial levels — to 2035 for the business-as-usual scenario and to 2040 for the low-emissions case (see 'Arctic methane'). This will lead to an extra \$60 trillion (net present value) of mean climate-change impacts for the scenario with no mitigation, or 15% of the mean total predicted cost of climate-change impacts (about \$400 trillion). In the low-emissions case, the mean net present value of global climate-change impacts is \$82 trillion without the methane release; with the pulse, an extra \$37 trillion, or 45% is added (see Supplementary Information). These costs remain the same irrespective of whether the methane emission is delayed by up to 20 years, kicking in at 2035 rather than 2015, or stretched out over two or three decades, rather than one. A pulse of 25 Gt of methane has half the impact of a 50 Gt pulse.

The economic consequences will be distributed around the globe, but the modelling shows that about 80% of them will occur in the poorer economies of Africa, Asia and South America. The extra methane magnifies flooding of low-lying areas, extreme heat stress, droughts and storms.

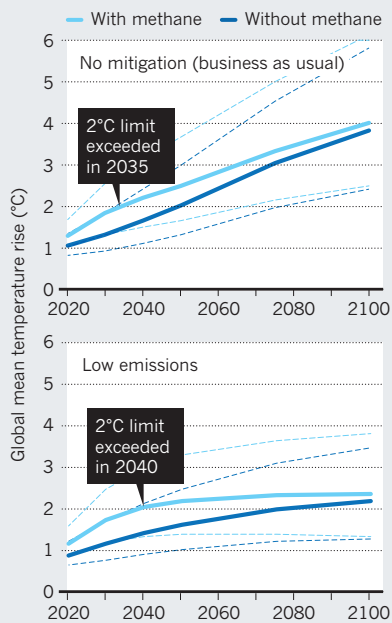
GLOBAL PROBLEM

The full impacts of a warming Arctic, including, for example, ocean acidification and altered ocean and atmospheric circulation, will be much greater than our cost estimate for methane release alone.

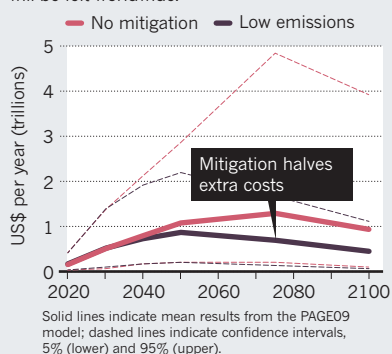
To find out the actual cost, better models are needed to incorporate feedbacks that

ARCTIC METHANE

Global mean temperatures will rise more quickly if 50 gigatonnes of methane is released from permafrost beneath the East Siberian Sea.



The economic impacts of the methane pulse will be felt worldwide.



are not included in PAGE09, such as linking the extent of Arctic ice to increases in Arctic mean temperature, global sea-level rise and ocean acidification, as well as including estimates of the economic costs and benefits of shipping. Oil-and-gas development in the Arctic should also, for example, take into account the impacts of black carbon, which absorbs solar radiation and speeds up ice melt, from shipping and gas flaring.

Splitting global economic impact figures into countries and industry sectors would raise awareness of specific risks, including the flooding of small-island states or coastal cities such as New York by rising seas. Mid-latitude economies such as those in Europe and the United States could be threatened, for example, by a suggested link between sea-ice retreat and the strength and position of the jet stream⁸, bringing extreme winter and spring weather. Unusual positioning of the jet stream over the Atlantic is thought to have caused

this year's protracted cold spell in Europe.

Such integrated analyses of Arctic change must enter global economic discussions. But neither the World Economic Forum (WEF) in its *Global Risk Report* nor the International Monetary Fund in its *World Economic Outlook*⁹ recognizes the potential economic threat from changes in the Arctic.

In 2012, noting that the far north is increasing in strategic importance and citing the need for informal dialogue among world leaders, the WEF launched its Global Agenda Council on the Arctic. This is welcome but more action is needed. The WEF should kick-start investment in rigorous economic modelling. It must ask world leaders to consider the economic time bomb beyond short-term gains from shipping and extraction.

The WEF should also encourage innovative adaptation and mitigation plans. It will be difficult — perhaps impossible — to avoid large methane releases in the East Siberian Sea without major reductions in global emissions of CO₂. Given that the methane originates in local seabed warming, then reducing black carbon deposits on snow and ice might buy some precious time¹⁰. But unknown factors could also mean that our impact estimates are conservative. Methane emerging in a sudden burst could linger for longer in the atmosphere, and trigger more rapid temperature changes than if the gas were released gradually.

Arctic science is a strategic asset for human economies, because the region drives critical effects in our biophysical, political and economic systems. Without this recognition, world leaders and economists will miss the big picture. ■

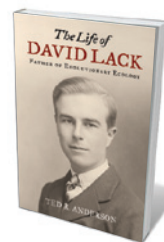
Gail Whiteman is professor of sustainability, management and climate change at Erasmus University Rotterdam, the Netherlands. **Chris Hope** is a reader in policy modelling at Judge Business School, University of Cambridge, UK.

Peter Wadhams is professor of ocean physics at the University of Cambridge, UK. e-mail: gwhiteman@rsm.nl

1. Gautier, D. L. *et al. Science* **324**, 1175–1179 (2009).
2. Smith, L. C. & Stephenson, S. R. *Proc. Natl Acad. Sci. USA* **110**, E1191–E1195 (2013).
3. Emmerson, C. & Lahn, G. *Arctic Opening: Opportunity and Risk in the High North* (Chatham House–Lloyd's, 2012); available at <http://go.nature.com/ruby4b>.
4. Wadhams, P. *AMBIO* **41**, 23–33 (2012).
5. Maslowski, W., Kinney, J. C., Higgins, M. & Roberts, A. *Annu. Rev. Earth Planet. Sci.* **40**, 625–654 (2012).
6. Shakhova, N. E., Alekseev, V. A. & Semiletov, I. P. *Doklady Earth Sci.* **430**, 190–193 (2010).
7. Hope, C. *Clim. Change* **117**, 531–543 (2013).
8. Francis, J. A. & Vavrus, S. J. *Geophys. Res. Lett.* **39**, L06801 (2012).
9. International Monetary Fund. *World Economic Outlook* (IMF, 2013).
10. Shindell, D. *et al. Science* **335**, 183–189 (2012).



David Lack (left) seeking out nests of chimney swifts in Ithaca, New York.



The Life of David Lack: Father of Evolutionary Ecology

TED R. ANDERSON
Oxford University
Press: 2013.

LACK FAMILY

to our understanding of a major issue in evolutionary biology by means of a few months' rather chaotic field study reflects a different age of scientific enquiry. So, too, does Lack's completely missing the point in his first major paper on the subject, in which he dismissed the possibility of adaptive causes of beak-size differences.

The central message of Anderson's book is that Lack should be understood as someone who bridged the gap between traditional natural history and the development of its modern academic descendant, evolutionary ecology. Indeed, Anderson argues that Lack's depth and breadth of impact were such that he deserves to be credited as a parent of the discipline.

Lack's scientific career began at a time when boundaries between amateur and professional were still blurred. He taught science at the progressive Dartington Hall School in Devon, and involved his pupils in early studies of robin behaviour. Mayr and evolutionary biologist Julian Huxley acted as mentors well before Lack established himself in his first academic post, as director of the Edward Grey Institute of Field Ornithology at Oxford University, UK (the post I now hold). Once in the Establishment, he withdrew from engaging with non-academic ecology. Anderson hints that this is most likely to have been due to conflicts with Max Nicholson — a giant of twentieth-century conservation and pioneer of what would now be called citizen science. They clashed over the academic direction of the institute, to which Lack was appointed against Nicholson's wishes.

Although he worked exclusively on birds, Lack's insights could frequently be generalized to other taxa. Birds are enduringly popular with macroecologists, thanks to the tremendous volume of information collected by amateur ornithologists, and (ironically) their rather limited diversity. Lack worked as the 'modern synthesis' in evolutionary biology was being developed (by Mayr among others). He was, if not one of its architects, one of its first practitioners.

Lack was particularly adept at using comparative approaches to derive general rules about how ecology affects evolution. He also provided much of the underpinning for modern population biology, using information from studies of marked individuals to estimate lifespans and rates of reproduction in the wild. Again, the foundations of this work were laid while Lack

EVOLUTIONARY BIOLOGY

Rare bird of evolution

Ben Sheldon applauds a life of the schoolmaster who broke new ground on Darwin's finches and speciation.

Charles Darwin had remarkably little to say about how the birds that bear his name — Darwin's finches — came to have such a variety of beaks, despite their iconic status in evolutionary biology. It was left to an English schoolmaster on sabbatical in the late 1930s to carry out the first serious work on this question.

David Lack's insight, gleaned from a four-month field trip to the Galapagos archipelago in 1938–39, ultimately centred on the role of ecological isolation in speciation. As Ted Anderson shows in his charming and very readable biography, Lack's argument, which developed slowly, was that differences in the way populations adapt to and compete for

local resources (such as seeds, in the case of finches) is a key part of the process of speciation. Although Lack completely changed his mind in print between 1940 and 1947, his work was extremely influential. Ernst Mayr, an evolutionary biologist with whom Lack corresponded extensively, credited Lack with introducing the idea of ecological isolation to the process of speciation. More recently, Canadian evolutionary biologist Dolph Schluter quipped that Darwin's finches, which have continued to provide a rich model for work on speciation, natural selection and evolutionary genetics, should be known as Lack's finches.

That a then schoolmaster contributed

was a schoolmaster, and are described in his classic *The Life of the Robin* (1943). At Oxford, he established population studies of the swift and the great tit, now in their seventh decade.

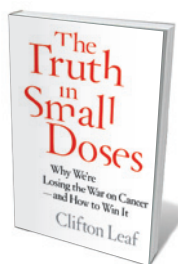
The Life of David Lack is organized chronologically around the 13 books Lack wrote before his death at the age of 62. This approach interweaves his science and life, but at times seems forced. Lack's major contributions to science arguably rest on just four of those books — *Darwin's Finches* and his three synthetic works on evolutionary ecology published between 1954 and 1968. A more in-depth exploration of the lasting influence of these four would be of considerable interest. Furthermore, the match between the content of the books when published and Lack's scientific and personal life at the time is not always very close.

However, Anderson gives us a vivid portrait of Lack and the personalities and careers of many people he interacted with. Lack was a socially reserved man, for whom family life was of great importance, but he was also often described as prickly and abrasive over scientific issues. Anderson does not shy away from discussing Lack's scientific and personal conflicts. For instance, he gives a clear exposition of the long-running debate between Lack and zoologist Vero Wynne-Edwards over the balance of group and individual-level selection in population regulation, which presages current debates on levels of selection. He takes us from the origin of the clash to an eventual rapprochement on a natural history trip in the Scottish Highlands.

The iciness between Lack and zoologist Charles Elton — a founder of population ecology who was a colleague of Lack's at Oxford and a neighbour for decades — remains something of a mystery, however. Despite sharing a building in Oxford's Botanic Garden, the door between Lack's and Elton's groups remained locked until Elton's retirement. Anderson speculates that the simple ecological principle of competitive exclusion might lie at the root of the conflict: here were two scientists who sought to explain the same general biological phenomena, using approaches that were too similar to enable their easy coexistence. Imagine how evolutionary ecology might have developed had they collaborated. ■

Ben C. Sheldon is Luc Hoffmann Professor of Field Ornithology and director of the Edward Grey Institute in the Department of Zoology at the University of Oxford, UK. e-mail: ben.sheldon@zoo.ox.ac.uk

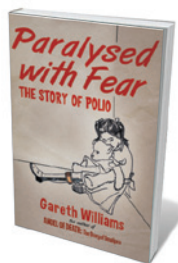
Books in brief



The Truth in Small Doses: Why We're Losing the War on Cancer — and How to Win It

Clifton Leaf SIMON & SCHUSTER (2013)

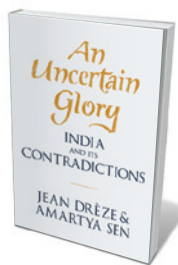
The US 'war on cancer' begun in 1971 has been an overall failure, argues journalist and cancer survivor Clifton Leaf. Over the past 40 years, he shows, crude deaths of US citizens from cancer have risen by 14%, although those from stroke and other killer diseases have fallen. The developing-world burden is also rising. In his exhaustively researched study probing why, Leaf points to a "cancer culture" in which scientists and medics think small, fail to coordinate results and focus on publishing rather than achieving breakthroughs.



Paralysed with Fear: The Story of Polio

Gareth Williams PALGRAVE MACMILLAN (2013)

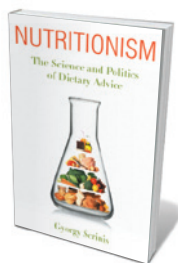
With the World Health Organization poised to roll out its Polio Eradication and Endgame Strategic Plan 2013–2018, the door could finally close on this devastating disease. Medical researcher Gareth Williams negotiates the hairpin bends of polio's history with aplomb. He takes us from its discovery by London medic Michael Underwood in the eighteenth century to Karl Landsteiner's isolation of the virus in 1908, and on through the twentieth century, when polio paralysed and killed millions, and consigned some to iron lungs or a life in callipers. A detailed, science-rich treatment.



An Uncertain Glory: India and its Contradictions

Jean Drèze and Amartya Sen ALLEN LANE (2013)

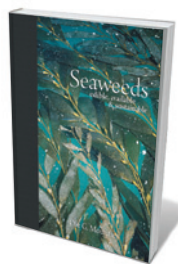
The world's largest democracy and one of its swiftest-growing economies lags behind many nations in immunization, education, medical care, the power sector and other key services. Economists Amartya Sen and Jean Drèze explore why India is "climbing up the ladder of per capita income while slipping down the slope of social indicators". China, for instance, contributes 2.7% of gross domestic product to public health; India, just 1.2%. A cogent synthesis of the state of a nation where high-tech success sits cheek by jowl with widespread open defecation and gross social inequality.



Nutritionism: The Science and Politics of Dietary Advice

Gyorgy Scrinis COLUMBIA UNIVERSITY PRESS (2013)

From diktats on salt to rulings on carbs, nutritional advice can turn supermarket aisles into minefields. Sociologist Gyorgy Scrinis blames "nutritionism", a reductive ideology that has dominated nutrition science for decades. It is a myth, he argues, that the interplay between nutrients, food and the body is fully understood. Meanwhile, much nutritional science focuses on individual nutrients such as fats, divorced from context such as overall diet. Scrinis calls for an integration of sound science with optimal production and processing, and hands-on cultivation and cooking.



Seaweeds: Edible, Available and Sustainable

Ole Mouritsen UNIVERSITY OF CHICAGO PRESS (2013)

Anyone who has wandered a wrack-strewn beach or munched nori-wrapped sushi knows the singular appeal of seaweeds. Biophysicist Ole Mouritsen trawls their biology and cultural roles as fertilizer, additives, medicine and food. Packed with minerals, proteins, trace elements and fatty acids, these algae are tasty, abundant and easily cultivable, and could feed future billions. Mouritsen even includes recipes: from seaweed pesto and dulse ice cream to kelp broth, a sea garden of delights. [Barbara Kiser](#)

was a schoolmaster, and are described in his classic *The Life of the Robin* (1943). At Oxford, he established population studies of the swift and the great tit, now in their seventh decade.

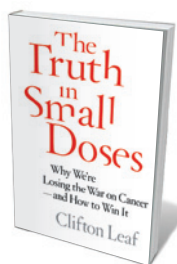
The Life of David Lack is organized chronologically around the 13 books Lack wrote before his death at the age of 62. This approach interweaves his science and life, but at times seems forced. Lack's major contributions to science arguably rest on just four of those books — *Darwin's Finches* and his three synthetic works on evolutionary ecology published between 1954 and 1968. A more in-depth exploration of the lasting influence of these four would be of considerable interest. Furthermore, the match between the content of the books when published and Lack's scientific and personal life at the time is not always very close.

However, Anderson gives us a vivid portrait of Lack and the personalities and careers of many people he interacted with. Lack was a socially reserved man, for whom family life was of great importance, but he was also often described as prickly and abrasive over scientific issues. Anderson does not shy away from discussing Lack's scientific and personal conflicts. For instance, he gives a clear exposition of the long-running debate between Lack and zoologist Vero Wynne-Edwards over the balance of group and individual-level selection in population regulation, which presages current debates on levels of selection. He takes us from the origin of the clash to an eventual rapprochement on a natural history trip in the Scottish Highlands.

The iciness between Lack and zoologist Charles Elton — a founder of population ecology who was a colleague of Lack's at Oxford and a neighbour for decades — remains something of a mystery, however. Despite sharing a building in Oxford's Botanic Garden, the door between Lack's and Elton's groups remained locked until Elton's retirement. Anderson speculates that the simple ecological principle of competitive exclusion might lie at the root of the conflict: here were two scientists who sought to explain the same general biological phenomena, using approaches that were too similar to enable their easy coexistence. Imagine how evolutionary ecology might have developed had they collaborated. ■

Ben C. Sheldon is Luc Hoffmann Professor of Field Ornithology and director of the Edward Grey Institute in the Department of Zoology at the University of Oxford, UK. e-mail: ben.sheldon@zoo.ox.ac.uk

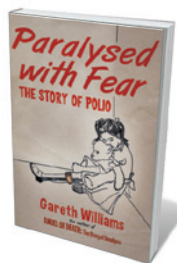
Books in brief



The Truth in Small Doses: Why We're Losing the War on Cancer — and How to Win It

Clifton Leaf SIMON & SCHUSTER (2013)

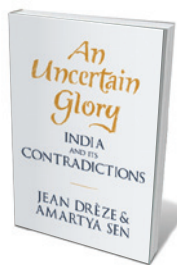
The US 'war on cancer' begun in 1971 has been an overall failure, argues journalist and cancer survivor Clifton Leaf. Over the past 40 years, he shows, crude deaths of US citizens from cancer have risen by 14%, although those from stroke and other killer diseases have fallen. The developing-world burden is also rising. In his exhaustively researched study probing why, Leaf points to a "cancer culture" in which scientists and medics think small, fail to coordinate results and focus on publishing rather than achieving breakthroughs.



Paralysed with Fear: The Story of Polio

Gareth Williams PALGRAVE MACMILLAN (2013)

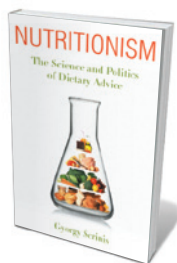
With the World Health Organization poised to roll out its Polio Eradication and Endgame Strategic Plan 2013–2018, the door could finally close on this devastating disease. Medical researcher Gareth Williams negotiates the hairpin bends of polio's history with aplomb. He takes us from its discovery by London medic Michael Underwood in the eighteenth century to Karl Landsteiner's isolation of the virus in 1908, and on through the twentieth century, when polio paralysed and killed millions, and consigned some to iron lungs or a life in callipers. A detailed, science-rich treatment.



An Uncertain Glory: India and its Contradictions

Jean Drèze and Amartya Sen ALLEN LANE (2013)

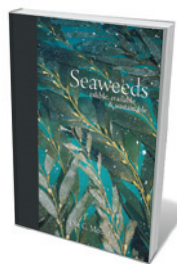
The world's largest democracy and one of its swiftest-growing economies lags behind many nations in immunization, education, medical care, the power sector and other key services. Economists Amartya Sen and Jean Drèze explore why India is "climbing up the ladder of per capita income while slipping down the slope of social indicators". China, for instance, contributes 2.7% of gross domestic product to public health; India, just 1.2%. A cogent synthesis of the state of a nation where high-tech success sits cheek by jowl with widespread open defecation and gross social inequality.



Nutritionism: The Science and Politics of Dietary Advice

Gyorgy Scrinis COLUMBIA UNIVERSITY PRESS (2013)

From diktats on salt to rulings on carbs, nutritional advice can turn supermarket aisles into minefields. Sociologist Gyorgy Scrinis blames "nutritionism", a reductive ideology that has dominated nutrition science for decades. It is a myth, he argues, that the interplay between nutrients, food and the body is fully understood. Meanwhile, much nutritional science focuses on individual nutrients such as fats, divorced from context such as overall diet. Scrinis calls for an integration of sound science with optimal production and processing, and hands-on cultivation and cooking.



Seaweeds: Edible, Available and Sustainable

Ole Mouritsen UNIVERSITY OF CHICAGO PRESS (2013)

Anyone who has wandered a wrack-strewn beach or munched nori-wrapped sushi knows the singular appeal of seaweeds. Biophysicist Ole Mouritsen trawls their biology and cultural roles as fertilizer, additives, medicine and food. Packed with minerals, proteins, trace elements and fatty acids, these algae are tasty, abundant and easily cultivable, and could feed future billions. Mouritsen even includes recipes: from seaweed pesto and dulse ice cream to kelp broth, a sea garden of delights. [Barbara Kiser](#)



A purported yeti footprint from the Menlung Basin in Nepal.

followed by other people faking evidence, until a self-reinforcing legend is established. As more monster hunters flock to find the beast, more dubious evidence is generated. Before you know it, there is a souvenir shop selling T-shirts.

Yet it is the hunters — ranging from outright rogues to serious, if misguided, researchers — who make this a gripping read. For instance, Bernard Heuvelmans, referenced as the founder of modern cryptozoology, earned a doctorate studying aardvark teeth, worked as a jazz musician and comedian, escaped from the Nazis and befriended Tintin creator Georges Prosper Remi (known by the pen-name ‘Hergé’) before producing his work on cryptids of all kinds.

On a 1958 expedition to Tibet to seek the yeti (a kind of Himalayan Bigfoot), led by Texas oil baron Tom Slick, some members of the group allegedly performed sleight of hand on a sacred relic, swapping human finger bones for purported yeti bones. The stolen bones were reportedly smuggled out of the country in the luggage of actor James Stewart.

Going by these and other anecdotes in *Abominable Science!* — such as people strapping on fake wooden feet to create ‘Sasquatch’ tracks — the book could justifiably have been a compilation of mockery and humour. In fact, it is a sensitive but devastating takedown of an entire subculture. It culminates in a final chapter that poses a puzzling question: why do people believe in monsters, in a world under surveillance by satellites and camera-toting hikers? No credible photograph of any of these creatures has ever been captured.

Rather unsatisfactorily, Loxton and Prothero fail to pin down the answer to this question. Instead, they end up detailing the damage pseudoscience can do, and suggesting how cryptid hunting could become a genuine science, contributing to “a world somewhat less ignorant and anti-scientific than the one in which we grew up”. On the basis of the information they present, however, it seems probable that true believers in such tall tales will always be with us. Evidence — or the lack of it — is never enough to scupper a good story. At least *Abominable Science!* proves that proper examination of it can produce an equally compelling tale.

Loxton admits that as a child, he believed in the things he now debunks. His passion eventually led him to question what he was told, and to co-author this book. Hopefully it will encourage others to take a similar journey. ■

Daniel Cressey is a reporter for Nature.

CRYPTOZOOLOGY

Beastly fakes

Daniel Cressey delves into a sceptics’ history of monster hunters and their mythical quarry.

The animals that star in *Abominable Science!* will be known to most children. Sadly, none of them exist. Daniel Loxton, a journalist for *Skeptical Science* magazine, and palaeontologist Donald Prothero dedicate their engaging book to chronicling how we arrived at a point at which, despite a huge absence of evidence, people still believe in cryptids such as the yeti, the Loch Ness monster, sea serpents and the ‘Congo dinosaur’ (Mokele Mbembe), a purported living sauropod.

What emerges is a never less than rigorous examination of the evidence, and a cultural history of cryptozoology. Loxton and Prothero, who take it in turns to pen chapters, have a grudging respect for some of the characters who populate this bizarre world of monster hunters, if more for their doggedness than their scientific precision. They point out that a number of real animals, such as the okapi and mountain gorilla, have been discovered after scientists investigated the stories of locals and vague reports of sightings. “The distinction between cryptozoology and conventional



Abominable Science! Origins of the Yeti, Nessie, and Other Famous Cryptids

DANIEL LOXTON AND DONALD R. PROTERO
Columbia University Press: 2013.

As Loxton and Prothero demonstrate, the hunters’ behaviour would appal most right-thinking field researchers. The title of the book says it all.

Each chapter follows a similar pattern, starting with an early pioneer combining some folk tales with a vague sighting of something in the distance. This is generally

biology is not necessarily very big,” they write, before comprehensively documenting that ‘necessarily’ is a key caveat.

Some hunters of Bigfoot, or Sasquatch, the gigantic primate that supposedly stalks the forests of North America, do seem to behave like biologists. A less kind interpretation would be that they adopt the trappings of science, but ignore its philosophy.



Observatories at sunrise on Haleakala, Hawaii.

Q&A Dane Maxwell Mountain guardian

Dane Maxwell is the specialist in Hawaiian traditional culture advising the US National Solar Observatory and its partners as they build the world's largest solar observatory atop Haleakala on the island of Maui. He talks about the sacred mountain, known as the house of the Sun, and local responses to the Advanced Technology Solar Telescope (ATST).



NICK HIGGINS, BASED ON A PHOTO BY DANE MAXWELL

How did you become interested in native Hawaiian cultural issues?

I grew up with my grandfather, who was involved in native Hawaiian politics; my grandmother was a hula teacher. I was sent to a native Hawaiian school. Everything I learned was around the Hawaiian culture, from hula to politics to prayers and other duties our ancestors used to have. My entire life was focused around what I do today.

What does it mean to be a kahu, or spiritual leader, like yourself?

Some people consider a kahu to be like a pastor, but the role is more like that of a Hawaiian shepherd. One thing I was trained in is our duty to human bones that have been dug up in construction projects. In Hawaiian

culture, your spiritual energy is in your bones. That's why reinterring bones is so important. The remains of the fire goddess Pele are said to be within Haleakala. The lava in that area is all her essence.

What's a typical day like for you, advising construction crews on the ATST?

Having grown up with deep respect for the land, it is really hard to watch excavation happening. They are digging out a really large pit and pouring in a lot of concrete. My concern is that people should show basic respect and reverence to the area. A lot of my time is spent making sure that construction workers stay on site and don't throw trash around.

There are a lot of these old rusted vehicles leaking hydraulic fluid everywhere, and you'll see these large stains on this beautiful area. Keeping it as natural and pristine as possible goes hand in hand with culture and nature. I'm up there almost every day, from

sunrise to about three-thirty in the afternoon, sometimes until sunset. It's beautiful. I get to see the mountain in sleet, snow, rain, wind and clear skies.

Many Hawaiians object to the ATST being built. Why are you working on it?

My grandfather and I opposed this project. But we really wanted to conserve as much as possible and do it in the right way. Otherwise it would be like turning our backs on our ancestors. It's our responsibility to take care of the mountain and preserve what we can.

Are cultural issues receiving more attention now than when the first telescopes were built on Haleakala in the 1950s and 60s?

Things are definitely getting better. For this project, the first thing we started with was the removal of a historical site, the Reber Circle, where a radio telescope had been. That's one of the highest points on the mountain, and in Hawaiian culture the highest points have the greatest significance. It was nice to remove that and clean up the mess.

A lot of the complaints are about how big and visible the ATST will be. What do you think about this?

It will be 14 storeys tall — really, really big. I tried to have them paint it another colour, but it had to be white because of its thermal properties. When the US Air Force built its telescope on Haleakala, it wanted to use materials to reflect the sky. But it reflected the Sun, and you can see it everywhere now.

What do you think astronomers ought to know about native Hawaiian concerns?

We don't hate astronomy. It's just that we want to manage the area better. We have an appreciation for science, and we want the scientists to have the same respect and reverence for the site. I always ask — if Hawaii was still its own independent state, would we have telescopes on Haleakala and Mauna Kea?

What does it mean to have a solar telescope on 'the house of the Sun'?

There are two hills up there. One of the demigods, called Maui, supposedly stood with one foot on one hill and one foot on the other hill. He reached up and lassoed the Sun and slowed it down in its passage. So the Sun has a lot of significance for Haleakala. If there has to be a telescope there, I guess it should be a solar telescope. A few of the locals support it. A lot of the community just wishes it could be a little shorter, or in another place on Earth. ■

INTERVIEW BY ALEXANDRA WITZE

Correspondence

Chocolate habits of Nobel prizewinners

There is a reported correlation between chocolate consumption by different nations and the number of Nobel laureates in their populations (F. H. Messerli *New Engl. J. Med.* 1562–1564; 2012). So what are the chocolate habits of Nobel prizewinners?

We surveyed 23 male winners of the Nobel prize in physics, chemistry, physiology or medicine, and economics. Ten (43%) reported eating chocolate more than twice a week, compared with only 25% of 237 well-educated age- and sex-matched controls ($P = 0.05$; see B. A. Golomb *et al. Arch. Intern. Med.* 172, 519–521; 2012). Three proffered that their chocolate consumption had indeed contributed to their Nobel prize, but most disavowed any link. Two attested that they had won the prize in spite of their chocolate habits (see also ‘Nobel laureates on chocolate’ and go.nature.com/copuha).

Some laureates might understate their chocolate consumption to amplify the impression that native prowess led to their prize, producing bias to the null. However,



given favourable chocolate–mortality associations, chocolate aficionados might have survived to get their just desserts.

Another caveat is that the laureates’ responses, like chocolate consumption itself, could be tongue in cheek.
Beatrice A. Golomb* *University of California, San Diego, California, USA.*
bgolomb@ucsd.edu

**On behalf of 14 co-authors (see go.nature.com/b5rz8h for a full list).*

Future food: politics plague seed banks

As a former head of the gene bank at the International Center for Tropical Agriculture (CIAT), I feel that Susan McCouch and others underestimate the political problems facing seed banks — particularly with respect to access to new samples (*Nature* 499, 23–24; 2013).

I disagree that the Convention on Biological Diversity “has created significant barriers to the sharing of genetic material”. Most of the banked seed samples that are shared internationally or used by plant breeders are beyond the control of the convention — including those from the institutes of the Consultative Group on International Agricultural Research (CGIAR) and the US Department of Agriculture.

Neither does the International Treaty on Plant Genetic Resources for Food and Agriculture “now govern access to crop diversity”. The United States, Russia, China, South Africa and 61 other countries have not ratified the treaty.

CGIAR holdings were included in the treaty on the understanding that they would remain freely available subject to a small tax (payable to the Food and Agriculture Organization of the

United Nations for the treaty’s benefit-sharing fund) on patented varieties derived from supplied samples.

The treaty also excludes some important crops, such as soya beans, groundnut, tomatoes, wild cassava, some wild wheat and maize (corn) varieties, sugar cane, oil palm and most fruits and nuts.

In my view, the treaty needs to include many more countries and be much more effective at ensuring access to seed samples, if it is to be a success.

Dave Wood Fyvie,
Aberdeenshire, UK.
agrobiodiversity@btinternet.com

Future food: use local knowledge

Efforts to mine seed banks and secure future food supplies would benefit from participation by farmers, who could help to develop crop varieties suitable for their own land-management systems (Susan McCouch *et al. Nature* 499, 23–24; 2013).

Farmers usually optimize land usage to supply year-round nutrition, to cater for cultural preferences and ecosystem services, to provide income, reduce labour and avoid economic risk.

Integrating these practices could enhance the success of crop-breeding efforts. In this age of information sharing, even remote farmers could now become directly involved in and take advantage of a globally accessible infrastructure of biodiversity informatics that incorporates local knowledge.

Such a biocultural informatics programme might add expense, but it would ultimately improve the cost-effectiveness of global food security and health.

Selena Ahmed *Tufts University, Medford, Massachusetts, USA.*
selena.ahmed@tufts.edu
Anna Herforth *Cornell University, Ithaca, New York, USA.*

COCOA CONNECTION

Nobel laureates on chocolate

“Laureates are handed chocolate replicas of the medal in gold foil at the Nobel ceremony. This must mean that the Nobel Foundation thinks chocolate is important.”

“My wife is a chocoholic but so far has not received the phone call.”

“Chocolate made me who I am today.”

“The correlation could stimulate even greater efforts

to compete with the Swiss, French and Belgians.”

“Your research is poised precisely in the vast space between balanced objectivity and total confusion.”

“I eat a lot of wine gums and biltong in case you are considering snacking in general.”

“I am upping my chocolate consumption to recapture my youthful IQ.”

Jerome Karle

(1918–2013)

Chemist who pioneered mathematical methods to solve crystal structures.

Shortly after the German physicist Max von Laue discovered in 1912 that crystals diffract X-rays, the British physicist Lawrence Bragg realized that he could use the pattern of X-rays diffracted from a crystal to pinpoint the precise positions of its constituent atoms. Yet as increasingly complex materials began to be studied, ever more clever methods were needed to 'solve' the crystal structures. Jerome Karle, Jerry to those who knew him, established new ways to convert an observed diffraction pattern into a set of atomic positions. He did so in principle by deriving mathematical formulae together with Herbert Hauptman, and he did so in practice by inventing procedures for applying the formulae to real crystals together with his wife, Isabella Karle.

The mathematical approaches that Karle and Hauptman established, known as direct methods, have helped researchers to elucidate the structure of key molecules such as vitamins and hormones, and to gain insight into biochemical mechanisms. Karle and Hauptman shared the 1985 Nobel Prize in Chemistry for their work.

Karle, who died on 6 June, was born in Brooklyn in New York city. A precocious product of New York public schools, he completed high school at just 15 years old and went on to the City College of New York. He graduated in 1937 along with Hauptman and Arthur Kornberg, another of City College's many Nobel laureates. He then went to Harvard University in Cambridge, Massachusetts, where he gained a master's degree in biology. After spending about a year at the New York State Health Department in Albany, Karle pursued further graduate studies, this time in chemistry at the University of Michigan in Ann Arbor. At Michigan, Karle studied the diffraction patterns resulting from firing electrons at gases. It was also here that he met Isabella Lugoski, a fellow graduate student, whom he married in 1942 and with whom he had three daughters.

After completing his dissertation in 1943, Karle moved to the University of Chicago to work on the Manhattan Project. He returned to Michigan in 1944 to take on a research project for the US Navy, which involved studying the structure of hydrocarbon lubricants. In 1946, he and his wife moved to the US Naval Research Laboratory (NRL) in Washington DC, where they remained until their retirement in 2009.

Initially, they continued to focus on



electron-diffraction experiments. In parallel, Karle made a theoretical analysis predicting what diffraction patterns to expect from oriented hydrocarbons, and this got him wondering about applying his theories to the analysis of crystal structures. It was around this point that Karle was joined by Hauptman.

The problem they faced was that although X-rays diffracted from crystals carry information that can produce a picture of the atomic structure, only part of that information is accessible experimentally. Only the amplitudes of the electromagnetic waves bouncing off the atoms can be observed by photon detectors; the phase offset of each periodic wave relative to the others cannot be measured. Fortunately, for typical crystals there are many more X-ray reflections than there are atoms, which implies that the reflections must be mathematically interrelated. Starting in 1950, Karle and Hauptman drew on fundamental knowledge about the nature of matter (specifically, that one cannot have negative electron density) to find mathematical relationships among the diffracted waves. Soon after, they established a probability theory, which they bravely announced in 1953 in an abstruse monograph entitled 'Solution of the Phase Problem'.

Early reception of the Karle–Hauptman work was at best muted. Quoting Karle himself, "during the early 1950s ... a large number of fellow-scientists did not believe a word we said." The tide was turned by Isabella applying the work to challenging structures such as peptides. In 1966, she and Karle published a landmark paper in *Acta*

Crystallographica, which laid out step by step how to determine crystal structures. Others joined the venture with computer programs, and ever increasing numbers of ever more complex structures came to be determined through direct methods.

By the time Karle and Hauptman received the Nobel prize, Karle had become prominent in crystallography circles, having served as President of the International Union of Crystallographers in the early 1980s.

As I discovered during my postdoctoral time with Karle in the early 1970s, the power of the statistical methods underlying his and Hauptman's approach is not unbounded (I tried with little success to apply his methods to protein crystals). Nevertheless, Karle's influence extends to macromolecules. He was fascinated by resonance in diffraction (whereby certain atoms behave anomalously when the energy of incident X-rays matches the energy of an electronic orbital), and he made seminal contributions to the theory underlying an approach now called multi-wavelength anomalous diffraction (MAD). MAD and SAD, MAD's single-wavelength counterpart, are now commonly used to determine macromolecular structures, such as membrane proteins. Both require that the resonant atoms be located as a first step, and the Karle–Hauptman direct methods are now the approach of choice for finding them.

Karle's interests were broad, as suggested by the name he gave his unit at the NRL — the Laboratory for the Structure of Matter. The work there ranged from electron diffraction of gases to quantum chemistry of excited states, to the study of glasses and amorphous materials, and of course, crystals. Although these activities engaged several group members and were largely experimental, the Jerry I knew was a lone theoretician; he authored many papers alone and his main working interaction was with a computer programmer who tested his theories.

Ultimately, Karle's major contribution was to allow researchers to shift their focus from the intricacies and challenges of crystallography to molecules and biochemical mechanisms. He turned chemical crystallographers into crystallographic chemists. ■

Wayne A. Hendrickson is at Columbia University, New York, USA. He was a postdoctoral fellow and then a research biophysicist with Karle from 1969 to 1984. e-mail: wayne@xtl.cumc.columbia.edu

NAVAL RES. LAB.

York, New York 10029, USA. **S. M.-F.** is also at the Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain. e-mails: mone.zaidi@mountsinai.org; simon.mendez-ferrer@cnic.es

1. Yang, W. *et al.* *Nature* **499**, 491–495 (2013).
2. Maes, C. *et al.* *Dev. Cell* **19**, 329–344 (2010).
3. Zaidi, M. *Nature Med.* **13**, 791–801 (2007).
4. Bowen, M. E. *et al.* *PLoS Genet.* **7**, e1002050 (2011).

5. Zaidi, M., Troen, B., Moonga, B. S. & Abe, E. J. *Bone Miner. Res.* **16**, 1747–1749 (2001).
6. Nakamura, T. *et al.* *Cell* **130**, 811–823 (2007).
7. Park, D. *et al.* *Cell Stem Cell* **10**, 259–272 (2012).
8. Zaidi, M., Sun, L. & Blair, H. C. *Cell Stem Cell* **10**, 233–234 (2012).
9. Mendez-Ferrer, S. *et al.* *Nature* **466**, 829–834 (2010).
10. Isern, J. *et al.* *Cell Rep.* **3**, 1714–1724 (2013).
11. Grossmann, K. S., Rosário, M., Birchmeier, C. & Birchmeier, W. *Adv. Cancer Res.* **106**, 53–89 (2010).
12. Kontaridis, M. I., Swanson, K. D., David, F. S., Barford, D. & Neel, B. G. *J. Biol. Chem.* **281**, 6785–6792 (2006).

nor most of the mass, which is in the cold molecular phase.

It is only in the past few years that we have had the capability to detect emissions from the mass-laden molecular phase of outflows in nearby galaxies. This is because of increases in sensitivity made possible by updated technology on telescopes such as the Very Large Array in New Mexico and the IRAM Plateau de Bure interferometer in France, and with the advent of Europe's Herschel Space Observatory and now the Atacama Large Millimeter Array (ALMA) in Chile. One of the surprising results of these studies is that massive molecular winds are much more common than previously thought. However, many unanswered questions remain, meaning that it will be essential to make detailed investigations of the nearest examples of these winds.

In their paper (page 450), Bolatto and colleagues¹ report on observations, made with ALMA during its first year of operations, of one of the nearest starburst galaxies with a well-known superwind, NGC 253. This allowed them to image the cold carbon monoxide (CO) gas emission in NGC 253 at an incredible resolution of 50 parsecs. The unprecedented sensitivity of the researchers' observations revealed a number of very faint CO emission streamers out to heights of 120–320 parsecs for the first time. The morphology of these filaments closely matches that of the known ionized gas wind, providing clear evidence that the outflow also contains a substantial molecular component.

The authors also found that these prominent molecular-gas features seem to be connected

ASTROPHYSICS

How to catch a galactic wind

Observations obtained by the Atacama Large Millimeter Array in Chile's Atacama Desert have revealed properties of the cold molecular phase of the wind that is being blown out of a nearby starburst galaxy. **SEE LETTER P.450**

MARK WESTMOQUETTE

In this issue, Bolatto *et al.*¹ present ground-breaking interferometric observations of carbon monoxide emission from the nucleus and the central 2-kiloparsec region of a nearby starburst galaxy. The high sensitivity and resolution of these observations allowed the authors to detect faint molecular gas associated with a known superwind and to measure its mass-outflow rate for the first time.

When galaxies interact or merge, large quantities of gas can sometimes be channelled inwards as a result of the galaxies' gravitational forces. Under the right circumstances, this can trigger a massive burst of star formation in the heart of either or both galaxies, known as a starburst. The energy released, or fed back, by all these young stars in the form of radiation and winds (fast-moving gas particles blown off the stars) can have a marked effect on the remaining gas in the vicinity.

A single, short-lived star-formation event in a galaxy has an effect similar to an explosion. The energy released by the new stars creates a shock wave that compresses the surrounding gas into an expanding bubble. If the bubble inflates to a size comparable to that of the galaxy itself, then it is referred to as a superbubble. However, if the starburst proceeds for an extended period, then the initial bubble inflates and bursts, and the ongoing injection of energy drives the gas into a continuous flow of material called a galactic wind or superwind².

The exact mechanisms of how the stars accelerate and drive gas outwards are hotly debated topics^{3,4}. Constraints on theoretical predictions can come only from observations, and, because galactic winds are composed of

materials of mostly very low density that range over about five orders of magnitude in temperature, meaningful observations are difficult to make. Until recently, most of our understanding of galactic winds has come from the warm (10,000 kelvin) ionized and cooler (5,000 K) neutral gas phases of the winds, because they emit or absorb light in the easily accessible optical part of the spectrum. However, these phases contain neither most of the energy, which is in the hot X-ray-emitting phase,

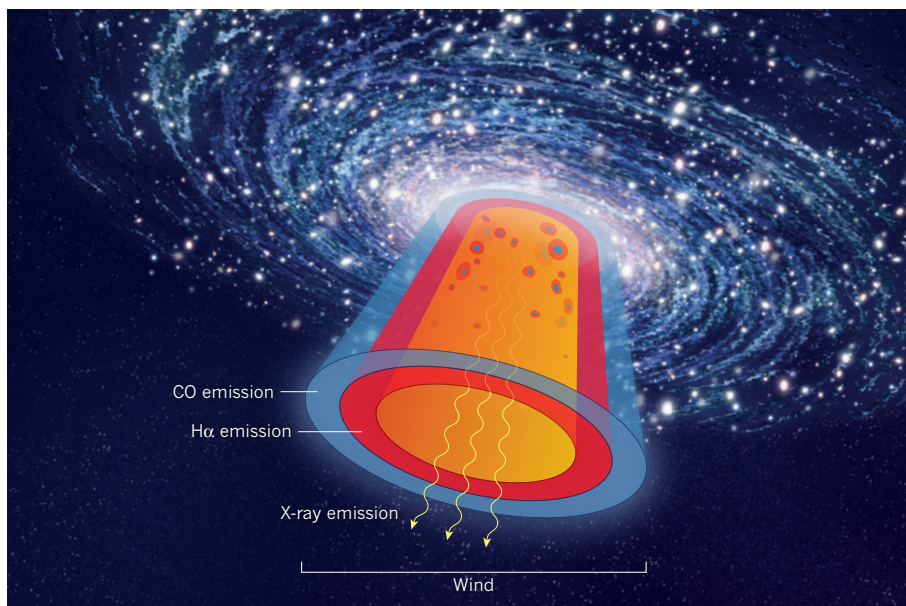


Figure 1 | The structure of the NGC 253 superwind. The wind being blown out of NGC 253 by the central starburst is thought to be approximately conical in structure with a temperature-stratified nature. The centre is filled with fast-moving and hot (a temperature of about 10^6 K) X-ray-emitting gas, and is surrounded by layers of cooler (10,000 K) $H\alpha$ -emitting gas and cold (100 K) molecular CO-emitting gas. Bolatto and colleagues' results¹ clearly show this cold CO layer at the edges of the warmer phases. The hot phase is also thought to contain clouds of cold material swept up by the wind⁶.

to the previously detected⁵ expanding molecular shells located on either side of the starburst region, thus providing an additional link between the starburst and the wind. The idea, built up from previous studies of winds, is that the hot wind fluid (temperatures greater than about 10^6 K), driven by the cumulative power of all the newly formed stars, interacts with the gas in the galaxy's disk and halo as it flows outwards. It is in the progressively cooler gas layers surrounding and embedded in the hot wind that the optical hydrogen- α (H α) emission (tracing the warm ionized gas at temperatures of about 10,000 K) and the molecular CO emission (tracing cold gas at temperatures of roughly 100 K) originates (Fig. 1).

The sensitivity of the authors' observations also allowed them to make direct measurements of the molecular-outflow speed and outflow rate. They found that the total mass-outflow rate is more than three times the measured star-formation rate. A considerable uncertainty exists in this measurement owing to the difficulties in converting from the measured outflowing CO mass to the predicted total outflowing mass, and in applying the correct geometric corrections (for example, for inclination), but it seems certain that the outflow rate is greater than the star-formation rate.

These results highlight several unanswered questions. The first relates to the depletion rate of gas and the starburst lifetime: if it is blowing out large amounts of gas, how long can the starburst continue given that it is removing the fuel for future star formation (that is, quenching itself)? Second, how is the energy required to eject this much material generated?

The mass-outflow rate sets constraints on the physical mechanisms needed to expel this material and thus forms important inputs for new theoretical models. Finally, how much of this ejected material is re-accreted later? One of the biggest unknowns with regard to starburst outflows is whether any of the ejected gas actually escapes the galaxy altogether. If it is not driven out far enough it may linger in the halo of the galaxy, cool, and rain back at a later time, potentially giving rise to a future starburst event, as might be happening in the starburst galaxy M82.

Although the molecular CO gas in the nuclear regions of NGC 253 has been imaged before⁵, the study by Bolatto *et al.* represents a major step forward in sensitivity and resolution, and provides new results on the wind's properties. It also showcases the revolutionary capabilities of ALMA in only its first year, operating with just 16 out of an eventual 66 antennas. ■

Mark Westmoquette is at the European Southern Observatory, 85748 Garching, Germany.

e-mail: westmoquette@gmail.com

1. Bolatto, A. D. *et al.* *Nature* **499**, 450–453 (2013).
2. Veilleux, S., Cecil, G. & Bland-Hawthorn, J. *Annu. Rev. Astron. Astrophys.* **43**, 769–826 (2005).
3. Murray, N., Ménard, B. & Thompson, T. A. *Astrophys. J.* **735**, 66 (2011).
4. Hopkins, P. F., Quataert, E. & Murray, N. *Mon. Not. R. Astron. Soc.* **421**, 3522–3537 (2012).
5. Sakamoto, K. *et al.* *Astrophys. J.* **636**, 685–697 (2006).
6. Strickland, D. K., Heckman, T. M., Weaver, K. A., Hoopes, C. G. & Dahlem, M. *Astrophys. J.* **568**, 689–716 (2002).

STRUCTURAL BIOLOGY

Meet the B family

The first crystal structures of class B G-protein-coupled receptors have been solved. They reveal features that might inform drug-development strategies for diseases ranging from osteoporosis to diabetes. SEE ARTICLES P.438 & P.444

PATRICK M. SEXTON & DENISE WOOTTEN

G-protein-coupled receptors (GPCRs) are the largest group of cell-surface receptors and are major targets for drug development^{1,2}. These proteins are characterized by a common architecture of seven transmembrane-spanning helical domains, and can be subdivided into three main groups: classes A, B and C. High-resolution structures of the membrane-spanning domain of GPCRs — the conduit for transmission of extracellular signals to the inside of a cell — provide snapshots that indicate how activating

and inactivating ligands modify the receptor structure. Until now, however, such studies have been principally restricted to class A receptors. In this issue, Hollenstein *et al.*³ (page 438) and Siu *et al.*⁴ (page 444) present the structures of the transmembrane domains of two class B members: corticotrophin-releasing factor-1 receptor and the glucagon receptor, respectively.*

Class B GPCRs include receptors for several peptide hormones, which are involved in a host of physiological functions from bone maintenance and glucose regulation to immune function and pain transmission. As a result, these receptors are targets for existing drugs that treat several disorders, including

*This article and the papers under discussion^{3,4} were published online on 17 July.

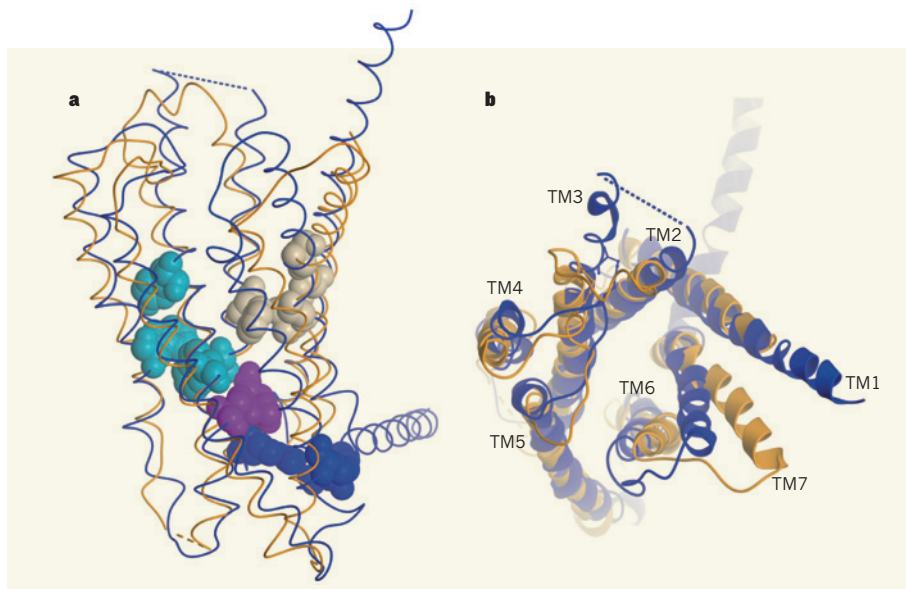


Figure 1 | Structural features of class B GPCRs. Hollenstein *et al.*³ and Siu *et al.*⁴ present the crystal structures of two class B G-protein-coupled receptors: CRF₁R (orange ribbons) and GCGR (blue ribbons), respectively. **a**, The structures reveal the locations of conserved amino-acid residues that form similar interactions in the two receptors, including between the transmembrane helices TM2, TM3 and TM4 (cyan), TM2 and TM3 (purple), TM1, TM2 and TM7 (beige), and TM2 and TM6 with the intracellular helix 8 (blue). **b**, The view of the proteins from outside the cell highlights the differences between the two structures at their extracellular faces, particularly in TM6 and TM7.

osteoporosis and type 2 diabetes, and are being actively pursued as targets for treating many more, from obesity and migraine to depression and chronic obstructive pulmonary disease.

Hollenstein and colleagues present a 3.0-ångström-resolution structure of the corticotrophin-releasing factor-1 receptor (CRF₁R) in complex with a small-molecule inhibitor. They arrived at this structure by introducing 12 thermostabilizing mutations into this GPCR and inserting the protein T4 lysozyme into its second intracellular loop. Siu and co-workers produced their 3.4-Å-resolution structure of the glucagon receptor (GCGR) using a version of the protein that was largely unmodified, except that its amino-terminal domain had been replaced with a thermally stabilized protein. The native N-terminal domain of class B GPCRs is crucial for peptide binding, but both teams removed this region to aid crystallization of the proteins.

As predicted, the core of both structures features seven transmembrane helices (TM1–TM7). However, although the relative positions of these helices at the intracellular face of the proteins overlap with those in class A GPCRs, there is substantial deviation between the two classes at the extracellular face. In both class B proteins, there are differences in the positioning of TM6 and TM7 that result in TM6 being shifted away from TM5, with TM1 seeming to move in parallel with TM7. This results in a wider and deeper extracellular cavity in the receptor core of the class B proteins that presumably forms part of the peptide-binding site. In addition, there are differences between the CRF₁R and GCGR structures themselves, in the upper segments of TM6 and

TM7 (Fig. 1). Although it is unclear whether these differences were influenced by the crystallization process, they indicate that the solution of transmembrane-core structures for other class B receptors will be required to help us understand how ligands bind and activate these proteins.

A major obstacle for the therapeutic targeting of class B receptors has been their notorious intractability for the identification of small-molecule ligands, in particular, small-molecule activators. The new structures shed light on why this is so: the openness of the receptors' binding pocket makes it difficult for a small ligand to engage sufficient key amino-acid residues to initiate activation of the receptor. Nonetheless, the solved structures show distinct subpockets that could represent sites for structure-based drug design.

Intriguingly, Hollenstein and colleagues' structure shows that the small-molecule inhibitor binds to a very deep pocket in the intracellular half of the CRF₁R core. This ligand forms extensive contacts with residues in TM3, TM5 and TM6, and presumably inhibits receptor activation by tethering the cytoplasmic half of TM6 to TM3 and TM5, thereby restricting conformational rearrangement of the intracellular face. This represents a new target for the design of small-molecule ligands. However, the amino-acid side chains in the equivalent region in the GCGR structure are more compact and would require reorganization to allow similarly sized ligands to bind.

The evolutionarily conserved amino-acid motifs in class A receptors have an important role in maintaining the receptors in an inactive (or weakly active) state. Although the

intracellular face of the class B receptors is similar to that seen for class A proteins (with the exception of an inward shift of TM7), some of the interactions that maintain the inactive class A conformations (including the ionic lock that tethers the cytoplasmic half of TM3 to TM6, the CWXP motif in TM6 and the NPXXY motif in TM7) are not present in the two class B receptors studied.

Class B receptors also have a distinct pattern of conserved amino-acid motifs that are important for maintenance of the inactive conformation and/or for conformational transitions required for activation. The CRF₁R and GCGR structures suggest conserved interactions between some of these key residues (Fig. 1). In addition, similar regions of contact are present between TM1 and TM2, TM1 and TM7, TM3 and TM4, and TM3 and TM6 in structures of both class A and B, although these interactions are mediated by different patterns of residues in each class. Thus, the new structures suggest that the two classes of proteins use distinct mechanisms for conformational control.

Although these reports represent a tremendous breakthrough in GPCR biology, as with all crystal structures, the intramembraneous class B structures provide only a snapshot of the receptors, which in reality are known (from cysteine-trapping studies⁵) to be highly dynamic proteins. Important questions remain about the final orientations of the N-terminal domains and transmembrane helices of the receptors, and about how natural activator molecules engage with both domains to activate the receptors. Answering these questions will require both crystallization of an intact ligand–receptor–G-protein complex and studies of receptor dynamics. ■

Patrick M. Sexton and Denise Wootten are at the Monash Institute of Pharmaceutical Sciences, Monash University, Parkville 3052, Victoria, Australia.

e-mails: patrick.sexton@monash.edu; denise.wootten@monash.edu

1. Fredriksson, R., Lagerström, M. C., Lundin, L. G. & Schiöth, H. B. *Mol. Pharmacol.* **63**, 1256–1272 (2003).
2. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. *Nature Rev. Drug Discov.* **5**, 993–996 (2006).
3. Hollenstein, K. *et al.* *Nature* **499**, 438–443 (2013).
4. Siu, F. Y. *et al.* *Nature* **499**, 444–449 (2013).
5. Dong, M. *et al.* *FASEB J.* **26**, 5092–5105 (2012).

CORRECTION

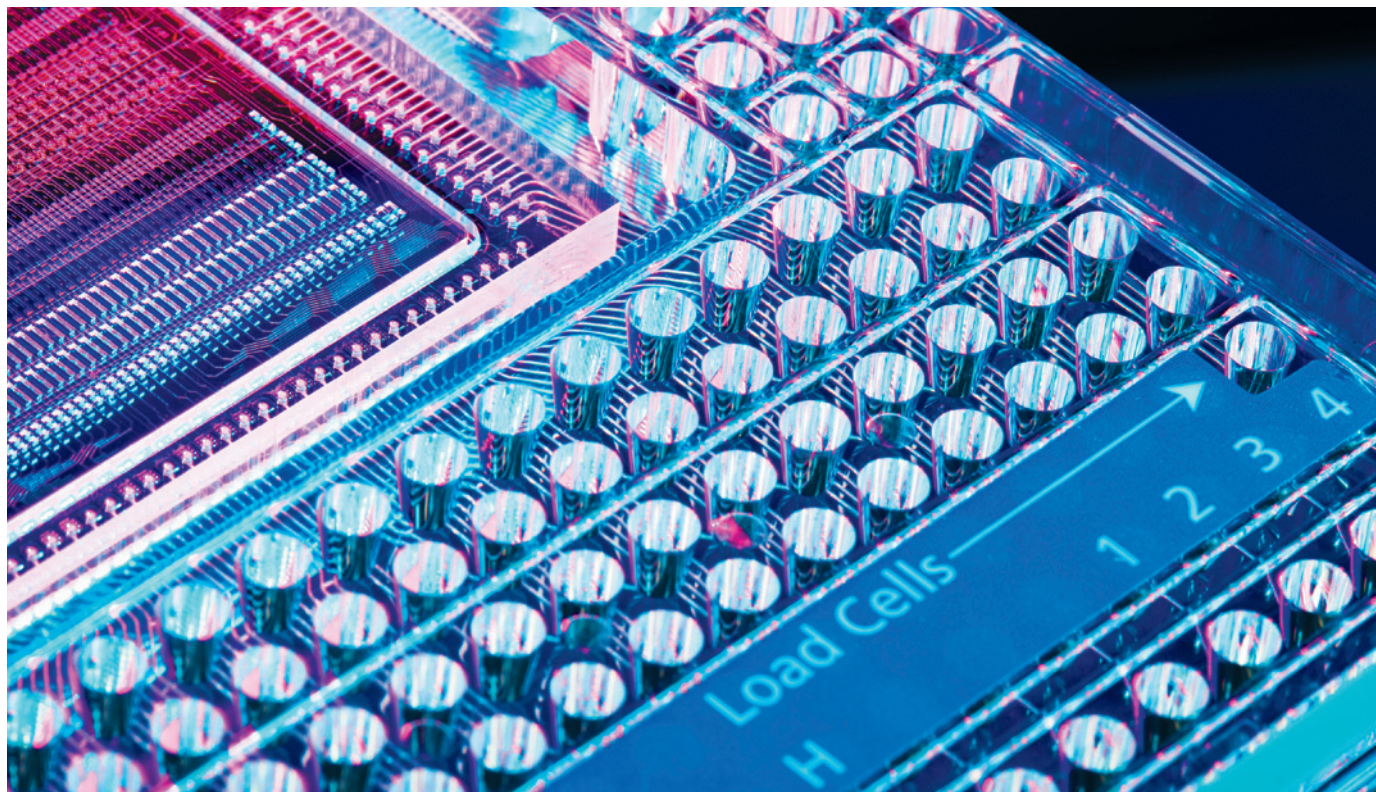
In the News & Views article 'Cloaking of heat' by Ulf Leonhardt (*Nature* **498**, 440–441; 2013), it was stated that "Einstein received his Nobel prize for his theory of diffusion, not for relativity." In fact, he received the prize "for his services to Theoretical Physics, and especially for his discovery of the law of the photoelectric effect".

TECHNOLOGY FEATURE

CANCER SHOWS STRENGTH THROUGH DIVERSITY

Tumours are made up of disparate cell populations that often resist treatment — but understanding this heterogeneity could provide ways to improve chemotherapy.

FLUIDIGM



Tiny channels in microfluidics chips such as this one from Fluidigm can isolate individual cancer cells for study.

BY CAITLIN SMITH

Cells come in all shapes and sizes — boxy epithelial cells, discoid red blood cells, delicate, threadlike neurons and the behemoth human egg that is just visible to the naked eye. Even among cells of the same basic type, no two are identical. And the same is true of cells within a cancerous tumour, where differences in size and shape can have profound implications for the progression of a patient's disease. As a result, researchers are keen to get to grips with cell heterogeneity. Developing the tools and techniques to rationalize this cellular chaos has been a slow process, but the latest methods for imaging, modelling and sorting cells

are at last coaxing them to relinquish their secrets.

For cancer, this may help to explain why a tumour that has been shrunk by chemotherapy suddenly kicks back into life and starts growing again. The plasticity of individual tumour cells lets them modify their behaviour in response to external cues, says Nicholas Saunders, a cancer biologist at the University of Queensland in Brisbane, Australia. One such cue is chemotherapy, and although the heterogeneity of tumour cells makes it harder to predict how each will respond to treatment, “we now have tools that allow us to interrogate this issue in a relatively definitive way,” he says. Recent techniques for sequencing the DNA of single cells from

tumours, for instance, has fired up this area as scientists explore ways to use the technology, says Saunders.

SINGLE LIFE

To investigate how cancer cells survive chemotherapy, researchers are moving into the challenging realm of single-cell analysis. At this small scale, it becomes hard to separate true variations between cells from technical errors in measurement, says Nicholas Navin, a molecular geneticist from the University of Texas MD Anderson Cancer Center in Houston. When differences between cells are detected, scientists can question whether the observed variations are important.

Researchers are particularly interested in ►

► the individual cells shed by tumours into a patient's bloodstream. Carried around the body, these 'metastatic' cells can initiate fresh tumours, allowing the disease to progress. But capturing these roaming cells for study is tricky, because they are mixed in with multiple cell types in the bloodstream.

One system Navin is using to isolate single tumour cells from blood is DEPArray, an instrument made by Silicon Biosystems of Bologna, Italy. This can isolate, move and image one tumour cell from a mixture of 100,000 cells.

Metastatic cells in the blood sample are first tagged with a chemical marker that emits light under a fluorescence microscope. In the DEPArray system, the individual cancer cells are then imprisoned in 'cages' created using an electric field. Viewed on a monitor, these cages can be manipulated to move a single cell into a collection vessel, ready for further study. The lack of physical contact helps the cells to stay alive during the manipulation.

Nevertheless, sorting takes time, says Navin. Initially, the DEPArray system took around an hour to isolate one tumour cell, but improvements to the technology and software mean that it can now move multiple tumour cells simultaneously from the mixture to the collection vessel. "The current system can route 13 cells in about 4 hours," says Navin.

Navin is also working with a system made by Fluidigm in South San Francisco, California. This captures 96 cells in one run, says Ken Livak, a researcher at the company. However, unlike DEPArray, it does not image cells to help with visual sorting, so it is best for isolating previously sorted cells, he says.

Fluidigm's system features a device about

the size of a postage stamp that contains tiny channels, valves and chambers. Minute amounts of fluid, along with cells, are driven through channels across the chip by opening and closing the valves. The channels contain a series of alcove-like capture sites. An unoccupied site will trap and hold an individual cell, but if the site is already occupied, the cells bypass it and move to the next one, until all 96 sites hold cells.

SHAPE-SHIFTERS

Chris Bakal, a cancer biologist at the Institute of Cancer Research in London, is hunting for patterns in the diversity of cancer-cell shapes. He and his team study metastatic melanoma cells, which are notorious for making drastic changes to their shape so that they can infiltrate far-flung reaches of the body.

The team's work centres on analysing images derived from spinning-disk confocal microscopy. In this technique, a laser illuminates the cells, and the microscope scans the light bouncing off them at many points simultaneously, gathered through pinholes in a spinning disk. The method is more sensitive than conventional confocal microscopy, which detects only one point of light at a time.

Bakal and his group have seen the diverse shapes of some cancer cells¹, and are now using statistical and computational analysis to try to identify which shapes are important.

"We think of these cells as extreme shape-shifters that can do anything," says Bakal. But generating and maintaining more diversity in cell shape than is needed may simply squander energy and drain the population of the cell shapes that are most useful, he says. Bakal and his team have found that, in fact, metastatic melanoma cells generally assume one of two shapes: rounded or spindle-shaped, each with its own advantages. "If you're a metastatic cell, you want those two shapes because a rounded

shape migrates through soft tissues like the brain or the circulatory system," he says, "whereas the spindle shape is good for bone and hard tissues."

Bakal thinks that looking at various aspects of heterogeneity in single cells will prove useful. "You might see genetic heterogeneity in this experiment, and you might see shape heterogeneity in another experiment", he says, but notes that it may not always be clear whether the two observations are connected. To determine if there is a link, he plans to sequence the DNA from individual cells after imaging them to see if he can find mutations that correlate with one shape or the other.

ORDER OUT OF CHAOS

Despite their heterogeneity, tumours cannot be totally chaotic, says Garry Nolan, a cell-signalling researcher at Stanford University in California. He thinks that there must be organization somewhere within the diversity, so his approach to studying individual cells focuses on differences in the patterns of the myriad proteins that cells express. He believes that the complement of proteins alters as a normal cell becomes cancerous. As a result, the different protein complements seen in a sample of cancer cells could be related to the past history of those cells.

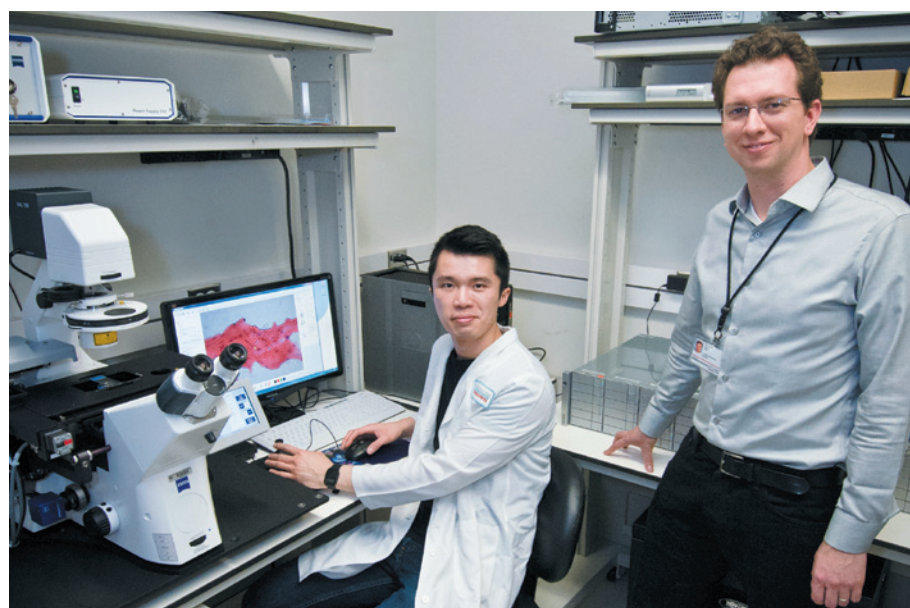
So far, Nolan's group has tracked more than 100 proteins simultaneously in individual cells, using a technique called mass cytometry. This is similar to flow cytometry, which separates cells according to fluorescently labelled proteins of interest. However, Nolan and his team wanted to look at many more proteins than is possible with flow cytometry, which is limited to the analysis of only a handful of proteins by the number of fluorescent tags that can be used. To solve this problem, the researchers developed mass cytometry so that they could identify tens or hundreds of proteins at the same time².

In mass cytometry, instead of proteins of interest being labelled with fluorescent markers, they are tagged with small metal particles that differ in mass. Once tagged, each cell is ionized and sent to a mass spectrometer, which separates the metal-tagged labels by mass. Unlike the fluorescent signals of flow cytometry, the mass measurements are relatively easy to distinguish from one another. Another benefit of this method is that it can measure proteins within the cell, because the cell is essentially vaporized during the process.

Nolan and his team are now developing their mass-cytometry technique to measure hundreds of proteins per cell, enabling them to piece together the puzzle of how cells become cancerous. The team has discovered a group of heterogeneous cancer cells that each have "their own little time-stamp signature on them", Nolan says. The varying complement of proteins on each cell indicates how far it has passed along the path to becoming cancerous, he adds.

By arranging the cancer cells according to these time-stamp proteins, the researchers

"We think of these cells as extreme shape-shifters that can do anything."



Marco Leung and Nicholas Navin (right) study genetic differences in cancer cells to gauge tumour activity.

created a timeline for a cell's physiology. Nolan believes that what seems to be a heterogeneous group of cancer cells is actually a snapshot of cells that represent different stages on a pathway leading to fully fledged cancer cells³. Viewed one by one, the mix may look wildly variable. But when viewed as a time-stamped group, "there is order there, waiting to be understood", Nolan says.

PERSONAL SPACE

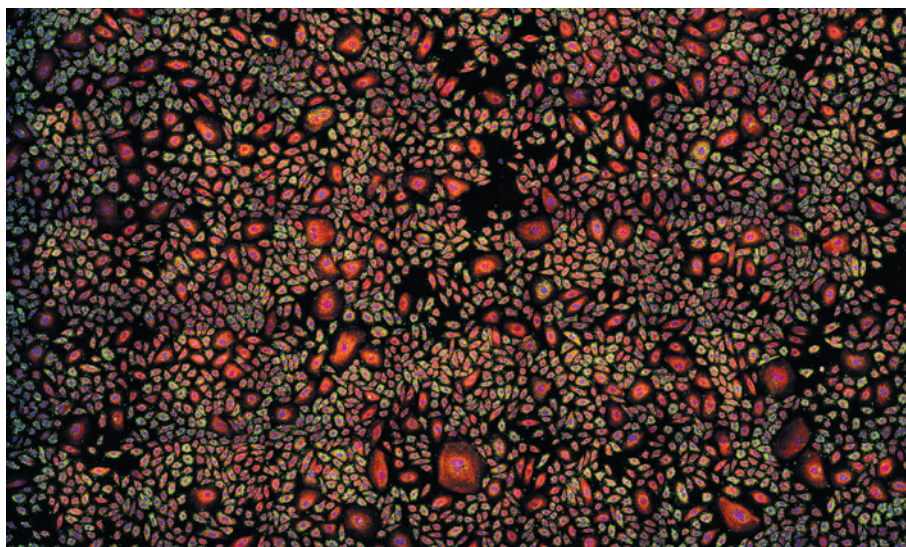
As well as presenting deviant shapes, cancer cells have a tendency to disregard the normal rules that other cells use for spacing themselves in three dimensions. "Basically they don't sit in these nice structures within the tissue like most cell types do," says Navin of breast-cancer cells. "They don't respect their neighbours."

The three-dimensional position of cancer cells within the tissue, as well as the areas immediately surrounding them, influence tumour formation and growth. Researchers hope that studying cellular heterogeneity in three dimensions — which more closely resembles real tissues — will deliver insights that could help to fight cancer.

Lucas Pelkmans, a researcher at the Institute of Molecular Life Sciences at the University of Zurich in Switzerland, studies how cells are affected by their surroundings. He and his team use automated, high-resolution imaging of millions of cells to monitor hundreds of parameters, including a cell's shape, distance from neighbouring cells and position within a tissue.

Pelkmans and his team then correlate these parameters with other measures of cellular activity, such as the molecular composition of cell membranes and the abundance of messenger RNA (mRNA) molecules, which are transcribed from DNA to serve as templates for the production of proteins in cells.

The team developed a technique that attaches fluorescent labels to single mRNA molecules of interest within individual cells, and then massively amplifies the fluorescent signal. "With that, you get a bright spot inside single cells," says Pelkmans. "By counting the



Using high-throughput fluorescence imaging of cells, a team at the University of Zurich is studying messenger RNA levels to understand the importance of a cell's spatial positioning within tissues.

number of spots, you basically get a read-out of the number of mRNAs in one cell." This can reveal whether a particular cell has different levels of gene expression compared to another; if it does, this might suggest that the two cells will go on to have different roles.

The varying types of correlation between the measured parameters create a tell-tale cell signature. "These signatures clearly can be different for different genes, but there are strong signatures," says Pelkmans. The signatures indicate a kind of tumour geography and hint at the functional role of a cancer cell at a given position. Cells can grow together as a community, but those on the periphery can show different signatures to those in the interior. Interpreting these signatures can help researchers to understand how signals exchanged between cells influence tumour growth.

The movement of cells within tumours has piqued the interest of Kornelia Polyak at the Dana-Farber Cancer Institute and Harvard Medical School in Boston, Massachusetts. She is studying the spatial changes that occur during cancer treatments. Using measurements from real tumour cells obtained from cancer patients, she and her team have built a computer model that simulates tumour growth.

The model allows researchers to take virtual samples of the simulated tumour at different times and places, Polyak says. They can even subject the simulated tumour to a course of cancer treatment. Although it is not yet ready for clinical use, Polyak hopes that the model will ultimately act as a surrogate patient, allowing clinicians to try out different simulated therapies and assess predicted outcomes before they treat patients.

"We could actually use this for designing the best treatment strategy," she says. "But the treatment itself changes the tumour, so you have to know how the tumour changes." Models might help physicians to get one step ahead

of the tumour, Polyak suggests, allowing them to anticipate the survival of a small population of drug-resistant cells and so quickly fight back against cancer recurrence⁴.

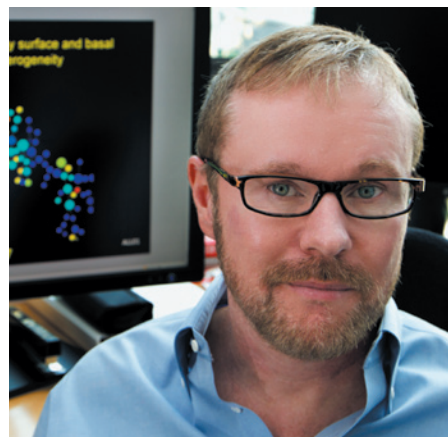
Key to the issue of resistance is knowing how many tumour cells already have the genetic mutations that make them resistant to drugs, and to what extent chemotherapy itself induces such mutations. Studying individual cells may provide the answer. "The more genetically diverse a tumour is, the more likely it is to be resistant to certain therapies, so that's one potentially useful parameter that you could get for patients," says Navin.

Understanding the degree of heterogeneity within a tumour is important in assessing the severity of the cancer. Individuals with diverse tumours might be more likely to harbour metastatic cells or be more resistant to therapy, compared with patients whose tumours are more homogeneous, says Navin. "If we can measure the extent of heterogeneity of a tumour-cell population, then we may be able to use this index to predict which patients will have invasive or metastatic tumours, and which will respond to chemotherapy or show poor survival," he says.

Cell heterogeneity gives normal cells the power to react to the environment, but it also underlies the ability of some tumour cells to emerge unscathed from even the strongest chemotherapy. If researchers can uncover how cancer cells adapt to cancer treatments, cell heterogeneity might ultimately be turned to the patients' advantage. ■

Caitlin Smith is a freelance writer in Portland, Oregon.

1. Yin, Z. *et al.* *Nature Cell Biol.* **15**, 860–871 (2013).
2. Bendall, S. C. *et al.* *Science* **332**, 687–696 (2011).
3. Bendall, S. C. & Nolan, G. P. *Nature Biotechnol.* **30**, 639–647 (2012).
4. Almendro, V., Marusyk, A. & Polyak, K. *Annu. Rev. Pathol.* **8**, 277–302 (2013).



Garry Nolan uses mass cytometry to track proteins that reveal a timeline of cancer-cell physiology.

LUCAS PELKMANS/UNIV. OF ZURICH

GARRY NOLAN/STANFORD UNIV.

SURRENDERED HUMAN

The path to redemption.

BY DEBORAH WALKER

I'm meditating in my room when Andrew bursts in (without knocking) and shouts out: "Guess what, Emma. There's a rumour that Mo-Sal will appear at the confirmation ceremony tomorrow."

"You've got to be kidding me." Mo-Sal: the commander of this colony. I begin to tremble.

"Just imagine," says Andrew, "breathing in his actual scent."

"It'd be overwhelming," I say. The synthetic amines that swirl around this lunar centre, the scent of ginger musk is so powerful. I shudder, as I imagine breathing in the real thing. "I don't know what to say."

"I know." Andrew flops onto my bed and stretches his arms above his head. His T-shirt rises up, exposing a swathe of taut stomach etched with a line of coarse blonde hairs. A small part of me acknowledges this, but most of me is thinking about Mo-Sal.

A thousand of us dressed in red, the colour of surrender. Today, we confirm our allegiance. Today, we will be assigned our roles. Tomorrow, we will be shipped out to a place in the Arium Domain.

The air is electric. All thoughts turned to Mo-Sal.

"Do you think he'll come?" asks Andrew. Andrew had been the comm officer on my ship. In another life we'd been close, and something of that still remained.

"I respect his decision," I say. Actually, I am willing him to come. To see him. To actually breathe in his scent. How wonderful that would be. I guess we're all thinking that, or most of us, at least. Humans are a poor species, driven by personal desires. I feel ashamed.

"It's all right, Em," says Andrew. "I understand."

I smiled. He was good to me. We were good together, once.

On the stage, the drums begin to beat. And we, we thousand surrendered supplicants, chant the words. I've been longing for this moment for six months, ever since my ship was captured. To think that I'd actually initiated the destruction sequence. It seems impossible that I'd been so resistant. We shouted the joyous words:

"We will surrender control to the Arium.

"We will accept our places in the community of the Arium Domain.

"We will be grateful."



Grateful. The quality of gratitude fills me, lifts me beyond my lower self. I am so grateful.

I walk up to the podium, when my name is called.

"Emma Catell, former captain of the Earth ship, *The Entangled Hope*."

I wince at the reminder of my previous life. To think I had been so arrogant as to seek to control the Arium. But there is no shame here. I know that because as I am holding out my hand to receive my assignment, something wondrous happens. A portal opens, a rip in space. Mo-Sal steps through the anomaly and onto the platform. His overwhelming scent. I sway on my feet. I see others falling to the floor. Mo-Sal is so powerful, but I have always had a quality of will, and I remain on my feet. I take no pride in that, it is a fault that led me to my previous actions, but all is forgotten here.

Mo-Sal lays his claw on my shoulder. I feel a wave of peace. I know that this is the defining moment of my life.

"You are Emma Catell." His dry-rasping words are resonant, portentous. He is all things. "You are the Earth captain responsible for one hundred and thirty-five Arium deaths."

I hang my head in shame.

"You are surrendered."

"I am, Mo-Sal."

"You have done well."

"Thank you," I whisper, remembering to receive a compliment graciously, but it is so difficult. I want to prostrate myself. I am an embarrassment.

"You will work on the surrendered Earth with the delinquents." Another reason for shame — some human have atrophied vomeronasal organs. They are unable to experience the chemical communion of the Arium. They are resisting the natural order of things.

"Thank you, Mo-Sal." This task is not what I hoped for. I would have liked to have served in the Arium Navy, bringing enlightenment to other worlds, but I gladly relinquish my personal desires and trust in Mo-Sal.

As Mo-Sal moves, a plume of his scent assaults me. I moan.

He touches me delicately on the chin. "Do not be ashamed, Emma Catell. You must embrace this." How kind he is. How much he understands.

I walk back to my place in the crowd. I glance at Andrew but his eyes and his thoughts are fixated on Mo-Sal, as mine should be.

As the ceremony proceeds, I think about Mo-Sal's words. I believe that he has given me a task. Species surrendered to the Arium often wither within a few generations. The desire for reproduction is subsumed in the love of the Arium. This is wrong. I want humans to serve the Domain for generations, stretching into the future. I will take the human approach in my new life. This is the task Mo-Sal has given me. I glance at Andrew.

The ceremony continues, until we're all assigned. Yet even in this moment of peace and wonder, there is a worm in my mind, an unwelcome remnant of my past self.

This is all biochemical, says the worm, whispering its vile thoughts. *This is control. Mere biological control.*

I am a great sinner, yet I'm filled with gratitude. This resistance inside of me is a gift. It means that I'll struggle for redemption. My subservience will be real and meaningful.

I whisper a prayer to the Arium for the opportunity to struggle in their honour. I breathe in the heady scent of Mo-Sal's musk. ■

Deborah Walker grew up in the most English town in the country, but she soon high-tailed it down to London where she now lives with her partner, Chris, and her two young children.

JACEY

Great ape genetic diversity and population history

Javier Prado-Martínez^{1*}, Peter H. Sudmant^{2*}, Jeffrey M. Kidd^{3,4}, Heng Li⁵, Joanna L. Kelley⁴, Belen Lorente-Galdos¹, Krishna R. Veeramah⁶, August E. Woerner⁶, Timothy D. O'Connor², Gabriel Santpere¹, Alexander Cagan⁷, Christoph Theunert⁷, Ferran Casals¹, Hafid Laayouni¹, Kasper Munch⁸, Asger Hobolth⁸, Anders E. Halager⁸, Maika Malig², Jessica Hernandez-Rodriguez¹, Irene Hernando-Herraez¹, Kay Prüfer⁷, Marc Pybus¹, Laurel Johnstone⁶, Michael Lachmann⁷, Can Alkan⁹, Dorina Twigg³, Natalia Petit¹, Carl Baker², Fereyoun Hormozdiari², Marcos Fernandez-Callejo¹, Marc Dabad¹, Michael L. Wilson¹⁰, Laurie Stevison¹¹, Cristina Camprubi¹², Tiago Carvalho¹, Aurora Ruiz-Herrera^{12,13}, Laura Vives², Marta Mele^{1†}, Teresa Abello¹⁴, Ivanela Kondova¹⁵, Ronald E. Bontrop¹⁵, Anne Pusey¹⁶, Felix Lankester^{17,18}, John A. Kiyang¹⁷, Richard A. Bergl¹⁹, Elizabeth Lonsdorf²⁰, Simon Myers²¹, Mario Ventura²², Pascal Gagneux²³, David Comas¹, Hans Siegmund²⁴, Julie Blanc²⁵, Lidia Agueda-Calpena²⁵, Marta Gut²⁵, Lucinda Fulton²⁶, Sarah A. Tishkoff²⁷, James C. Mullikin²⁸, Richard K. Wilson²⁶, Ivo G. Gut²⁵, Mary Katherine Gonder²⁹, Oliver A. Ryder³⁰, Beatrice H. Hahn³¹, Arcadi Navarro^{1,32,33}, Joshua M. Akey², Jaume Bertranpetit¹, David Reich⁵, Thomas Mailund⁸, Mikkel H. Schierup^{8,34}, Christina Hvilsom^{24,35}, Aida M. Andrés⁷, Jeffrey D. Wall¹¹, Carlos D. Bustamante⁴, Michael F. Hammer⁶, Evan E. Eichler^{2,36} & Tomas Marques-Bonet^{1,33}

Most great ape genetic variation remains uncharacterized^{1,2}; however, its study is critical for understanding population history^{3–6}, recombination⁷, selection⁸ and susceptibility to disease^{9,10}. Here we sequence to high coverage a total of 79 wild- and captive-born individuals representing all six great ape species and seven subspecies and report 88.8 million single nucleotide polymorphisms. Our analysis provides support for genetically distinct populations within each species, signals of gene flow, and the split of common chimpanzees into two distinct groups: Nigeria–Cameroon/western and central/eastern populations. We find extensive inbreeding in almost all wild populations, with eastern gorillas being the most extreme. Inferred effective population sizes have varied radically over time in different lineages and this appears to have a profound effect on the genetic diversity at, or close to, genes in almost all species. We discover and assign 1,982 loss-of-function variants throughout the human and great ape lineages, determining that the rate of gene loss has not been different in the human branch compared to other internal branches in the great ape phylogeny. This comprehensive catalogue of great ape genome diversity provides a framework for understanding evolution and a resource for more effective management of wild and captive great ape populations.

We sequenced great ape genomes to a mean of 25-fold coverage per individual (Table 1, Supplementary Information and Supplementary Table 1) sampling natural diversity by selecting captive individuals of known wild-born origin as well as individuals from protected areas in Africa (Fig. 1a). We also included nine human genomes—three African and six non-African individuals¹¹. Variants were called using the software package GATK (ref. 12) (Methods), applying several quality

filters, including conservative allele balance filters, and requiring that genomes showed <2% contamination between samples (Methods and Supplementary Information). In order to assess the quality of single nucleotide variant (SNV) calls, we performed three sets of independent validation experiments with concordance rates ranging from 86% to 99% depending on allele frequency, the great ape population analysed and the species reference genome used (Supplementary Information and Supplementary Table 2). In total, we discovered 84.0 million fixed substitutions and 88.8 million segregating sites of high quality (Table 1 and Supplementary Table 3), providing the most comprehensive catalogue of great ape genetic diversity to date. From these variants we also constructed a list of potentially ancestry-informative markers (AIMs) for each of the surveyed populations, although a larger sampling of some subspecies is still required (Supplementary Information).

We initially explored the genetic relationships between individuals by constructing neighbour-joining phylogenetic trees from both autosomal and mitochondrial genomes (Supplementary Information). The autosomal tree identified separate monophyletic groupings for each species or subspecies designation (Supplementary Fig. 8.5.1) and supports a split of extant chimpanzees into two groups. Nigeria–Cameroon and western chimpanzees form a monophyletic clade (>97% of all autosomal trees); central and eastern chimpanzees form a second group (72% of all autosomal trees).

Genome-wide patterns of heterozygosity (Fig. 1b) reveal a threefold range in single nucleotide polymorphism (SNP) diversity. Non-African humans, eastern lowland gorillas, bonobos and western chimpanzees show the lowest genetic diversity ($\sim 0.8 \times 10^{-3}$ heterozygotes per base pair (bp)). In contrast, central chimpanzees, western lowland gorillas

¹Institut de Biologia Evolutiva, (CSIC-Universitat Pompeu Fabra), PRBB, Doctor Aiguader 88, Barcelona, Catalonia 08003, Spain. ²Department of Genome Sciences, University of Washington, 3720 15th Avenue NE, Seattle, Washington 98195, USA. ³Department of Human Genetics, University of Michigan, 1241 E. Catherine Street, Ann Arbor, Michigan 48109, USA. ⁴Department of Genetics, Stanford University, 300 Pasteur Drive, Lane L301, Stanford, California 94305, USA. ⁵Department of Genetics, Harvard Medical School, Boston, 77 Avenue Louis Pasteur, Massachusetts 02115, USA. ⁶Arizona Research Laboratories, Division of Biotechnology, University of Arizona, 1041 E. Lowell Street, Tucson, Arizona 85721, USA. ⁷Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig, 04103, Germany. ⁸Bioinformatics Research Centre, Aarhus University, DK-8000 Aarhus C, Denmark. ⁹Bilkent University, Faculty of Engineering, Ankara, 06800, Turkey. ¹⁰Department of Anthropology, University of Minnesota, Minneapolis, Minnesota 55455, USA. ¹¹Institute for Human Genetics, University of California San Francisco, San Francisco, California 94143, USA. ¹²Departament de Biologia Cel·lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Catalonia 08193, Spain. ¹³Institut de Biociències i de Biomedicina, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Catalonia 08193, Spain. ¹⁴Parc Zoològic de Barcelona, Barcelona, Catalonia 08003, Spain. ¹⁵Biomedical Primate Research Centre, P.O. Box 3306, 2280 GH, Rijswijk, The Netherlands. ¹⁶Department of Evolutionary Anthropology, Duke University, Durham, North Carolina 27708, USA. ¹⁷Limbe Wildlife Centre, BP 878, Limbe, Cameroon. ¹⁸Paul G. Allen School for Global Animal Health, Washington State University, Washington 99164, USA. ¹⁹North Carolina Zoological Park, Asheboro, North Carolina 27205, USA. ²⁰Department of Psychology, Franklin and Marshall College, Lancaster, Pennsylvania 17604, USA. ²¹Department of Statistics, Oxford University, 1 South Parks Road, Oxford OX1 3TG, UK. ²²Department of Genetics and Microbiology, University of Bari, Bari 70126, Italy. ²³Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, California 92093, USA. ²⁴Department of Biology, Bioinformatics, University of Copenhagen, Copenhagen 2200, Denmark. ²⁵Centro Nacional de Análisis Genómico (CNAG), PCB, Barcelona, Catalonia 08028, Spain. ²⁶Genome Sequencing Center, Washington University School of Medicine, St Louis, Missouri 63108, USA. ²⁷Department of Biology and Genetics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ²⁸National Institutes of Health Intramural Sequencing Center (NISC), Bethesda, Maryland 20892, USA. ²⁹Biological Sciences, University at Albany, State University of New York, Albany, New York 12222, USA. ³⁰Genetics Division, San Diego Zoo's Institute for Conservation Research, 15600 San Pasqual Valley Road, Escondido, California 92027, USA. ³¹Departments of Medicine and Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ³²Instituto Nacional de Bioinformática, UPF, Barcelona, Catalonia 08003, Spain. ³³Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia 08010, Spain. ³⁴Department of Bioscience, Aarhus University, DK-8000 Aarhus C, Denmark. ³⁵Copenhagen Zoo, DK 2000 Frederiksberg, Denmark. ³⁶Howard Hughes Medical Institute, 3720 15th Avenue NE, Seattle, Washington 98195, USA. †Present address: Centre for Genomic Regulation (CRG), C/Dr Aiguader, 88, 08003 Barcelona, Spain.

*These authors contributed equally to this work.

Table 1 | Genetic variation summary by species and subspecies

Genus	Scientific name species/subspecies	Common name	N	Mean coverage	Fixed sites to human reference	No. of SNVs*	Mean SNVs per individual*	No. of singletons†	Ancestry informative markers (AIMs)‡	N_e (10^{-3})§
<i>Homo</i>	<i>Homo sapiens</i>	Non-African	6	18.3	386,974	5,887,443	2,639,546	1,379,448	12,316	9.7–19.5
		African	3	20.9	632,253	6,309,453	3,203,178	2,448,454	12,316	13.9–27.9
		Humans	9	19.2	224,660	9,172,573	3,061,604	3,827,902	NA	13.1–16.2
<i>Pan</i>	<i>Pan troglodytes</i>	Nigeria–Cameroon	10	16.7	25,017,403	12,605,585	4,816,435	2,695,109	2,213	18.5–37.0
		Eastern	6	28.7	25,126,506	11,264,879	4,843,530	2,228,396	1,265	19.7–39.5
		Central	4	23.8	25,080,750	11,820,858	4,983,933	3,948,347	619	24.4–48.7
		Western	4	27.3	26,832,247	4,729,933	2,411,501	1,481,079	145,548	9.8–19.5
		Common Chimpanzees	24	22.5	24,087,088	27,153,659	5,693,903	10,352,931	149,645	30.9–61.8
<i>Pan</i>	<i>Pan paniscus</i>	Bonobos	13	27.5	27,068,299	8,950,002	2,738,755	3,159,889	NA	11.9–23.8
<i>Gorilla</i>	<i>Gorilla beringei</i>	Eastern lowland	3	22.8	34,537,496	3,866,117	2,578,328	484,482	317,028	12.2–24.3
		Cross river	1	17.6	35,553,861	2,585,360	2,585,360	165,482	35,693	14.9–29.8
		Western lowland	23	17.8	31,602,620	17,314,403	6,410,662	2,797,388	19,902	26.8–53.5
		Gorillas	27	18.3	31,376,203	19,177,989	6,492,831	3,447,352	372,623	28.4–56.9
<i>Pongo</i>	<i>Pongo abellii</i>	Sumatran	5	28.7	62,880,923	14,543,573	7,263,256	5,681,303	1,132,808	27.5–55.0
		Bornean	5	25.8	64,249,235	10,321,213	5,763,354	3,555,596	1,132,808	19.5–39.0
		Orangutans	10	27.3	60,661,869	24,309,920	9,338,148	6,409,648	NA	42.3–84.6
		All	83	23.0	83,954,672	88,764,143	NA	NA	NA	NA

*Polymorphic variants found in each species/subspecies after subtracting fixed sites.

†Singletons and doubletons calculated combining all the samples within the species.

‡Variants only found in a single group within each species.

§Calculated from Θ_w : $\mu = 1 \times 10^{-9}$ to 0.5×10^{-9} mut bp⁻¹ yr⁻¹ and $g = 25$ for *Homo* and *Pan*, 19 for *Gorilla* and 26 for *Pongo*.

||Hybrid sample Donald and 4 related gorillas were excluded.

The combined data for groups is shown in bold.

NA, not applicable.

and both orangutan species show the greatest genetic diversity (1.6×10^{-3} – 2.4×10^{-3} heterozygotes per bp). These differences are also reflected by measures of inbreeding from runs of homozygosity¹³ (Fig. 1c and Supplementary Information). Bonobos and western lowland gorillas, for example, have similar distributions of tracts of homozygosity as human populations that have experienced strong genetic bottlenecks (Karitiana and Papuan). Eastern lowland gorillas appear to represent the most inbred population, with evidence that they have been subjected to both recent and ancient inbreeding.

To examine the level of genetic differentiation between individuals we performed a principal component analysis (PCA) of SNP genotypes (Supplementary Information). Chimpanzees were stratified between subspecies with PC1 separating western and Nigeria–Cameroon chimpanzees from the eastern and central chimpanzees and PC2 separating western and Nigeria–Cameroon chimpanzees. In gorillas, PC1 clearly separates eastern and western gorillas, whereas the western lowland gorillas are distributed along a gradient of PC2, with individuals from the Congo and western Cameroon positioning in opposite directions along the axis. The isolated Cross River gorilla is genetically more similar to Cameroon western lowland gorillas and can be clearly differentiated with PC3 (Supplementary Fig. 8.2.9).

We explored the level of shared ancestry among individuals within each group¹⁴ using an admixture model (FRAPPE). In chimpanzees, the four known subspecies are clearly distinguished when fitting the model using four ancestry components ($K = 4$) (Fig. 1d). Additional substructure is identified among the eastern chimpanzees Vincent and Andromeda ($K = 6$), who hail from the most eastern sample site (Gombe National Park, Tanzania). As in Gonder *et al.*², we have identified three Nigeria–Cameroon samples (Julie, Tobi and Banyo, $K = 3$ –5) with components of central chimpanzee ancestry. However, taking central chimpanzees and the remaining Nigeria–Cameroon chimpanzees as ancestral populations shows no evidence of gene flow by either the F3 statistic or HapMix. This indicates that these three samples are not the result of a recent admixture and may represent a genetically distinct population (Supplementary Information).

In gorillas, following the separation of eastern and western lowland species ($K = 2$), an increasing number of components further subdivide western lowland populations distinguishing Congolese and Cameroonian gorillas—a pattern consistent with the structure observed in the PCA analysis (Supplementary Fig. 8.2.9). One striking observation is the extent of admixed ancestry predicted for captive individuals when

compared to wild-born. Our analysis suggests that most captive individuals included in this study are admixed from two or more genetically distinct wild-born populations leading to an erosion of phylogeographic signal. This finding is consistent with microsatellite analyses of captive gorillas¹⁵ and the fact that great ape breeding programs have not been managed at the subspecies level.

As great apes have been evolving on separate lineages since the middle Miocene, we attempted to reconstruct the history of these various species and subspecies by applying methods sensitive to branching processes, changes in effective population size (N_e), and gene flow occurring at different time scales. Using a combination of speciation times inferred from a haploid pairwise sequential Markovian coalescent (PSMC) analysis¹⁶, a coalescent hidden Markov model (CoalHMM)³ and incomplete lineage sorting approaches, we were able to estimate the most ancient split times and effective population sizes among the great ape species. By combining these estimates with an approximate Bayesian computation (ABC)¹⁷ analysis applied to the more complex chimpanzee phylogeny, we constructed a composite model of great ape population history over the last ~15 million years (Fig. 2). This model presents a complete overview of great ape divergence and speciation events in the context of historical effective population sizes.

PSMC analyses of historical N_e (Fig. 3) suggests that the ancestral *Pan* lineage had the largest effective population size of all lineages >3 million years ago (Myr), after which time the population of the common ancestor of both bonobos and chimpanzees experienced a dramatic decline. Both PSMC and ABC analyses support a model of subsequent increase in chimpanzee N_e starting ~1 Myr, before their divergence into separate subspecies. Following an eastern chimpanzee increase in N_e (~500 thousand years ago, kyr), the central chimpanzees reached their zenith ~200–300 kyr followed by the western chimpanzee ~150 kyr. Although the PSMC profiles of the two subspecies within each of the major chimpanzee clades (eastern/central and Nigeria–Cameroon/western) closely shadow each other between 100 kyr and 1 Myr, the western chimpanzee PSMC profile is notable for its initial separation from that of the other chimpanzees, followed by its sudden rise and decline (Fig. 3 and Supplementary Information). The different gorilla species also show variable demographic histories over the past ~200 kyr. Eastern lowland gorillas have the smallest historical N_e , consistent with smaller present-day populations and a history of inbreeding (Fig. 1c). A comparison of effective population sizes with the ratio of non-synonymous to synonymous substitutions finds that selection has

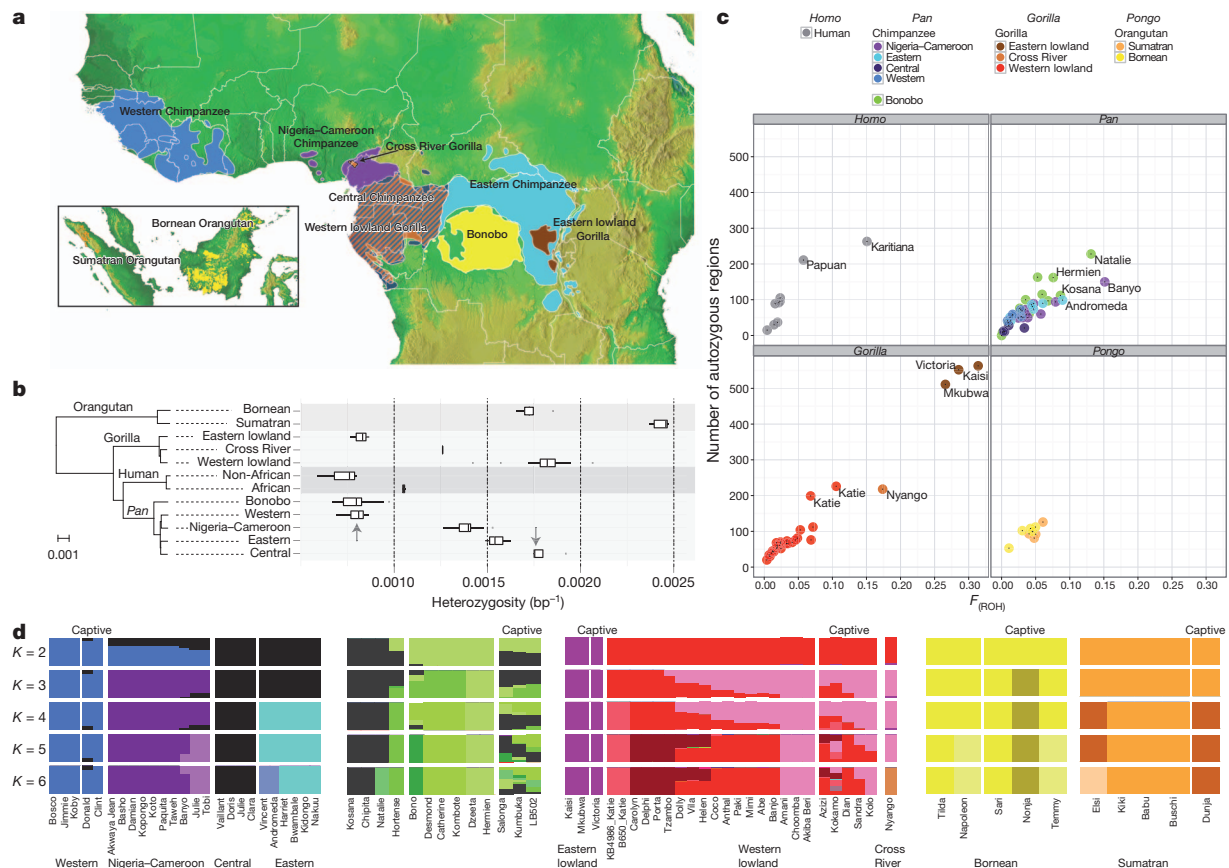


Figure 1 | Samples, heterozygosity and genetic diversity. **a**, Geographical distribution of great ape populations across Indonesia and Africa sequenced in this study. The formation of the islands of Borneo and Sumatra resulted in the speciation of the two corresponding orangutan populations. The Sanaga River forms a natural boundary between Nigeria–Cameroon and central chimpanzee populations whereas the Congo River separates the bonobo population from the central and eastern chimpanzees. Eastern lowland and western lowland gorillas are both separated by a large geographical distance. **b**, Heterozygosity estimates of each of the individual species and subspecies are superimposed onto a neighbour-joining tree from genome-wide genetic distance estimates (branch lengths in units of substitutions per bp). Arrows indicate heterozygosities previously reported³⁰ for western and central chimpanzee populations. **c**, Runs of homozygosity among great apes. The relationship

between the coefficient of inbreeding (F_{ROH}) and the number of autozygous >1 megabase segments is shown. Bonobos and eastern lowland gorillas show an excess of inbreeding compared to the other great apes, suggesting small population sizes or a fragmented population. **d**, Genetic structure based on clustering of great apes. All individuals (columns) are grouped into different clusters ($K = 2$ to $K = 6$, rows) coloured by species and according to their common genetic structure. Most captive individuals, labelled on top, show a complex admixture from different wild populations. A signature of admixture, for example, is clearly observed in the known hybrid Donald, a second-generation captive predicted to be a 15% admixture of central chimpanzee and a western background consistent with its pedigree. A grey line at the bottom denotes new groups at $K = 6$ in agreement with the location of origin or ancestral admixture.

acted more efficiently in populations with higher N_e , consistent with neutral theory (Supplementary Information).

Although the phylogeny of bonobos and western, central and eastern common chimpanzees has been well established based on genetic data¹⁸, there is still uncertainty regarding their relationship to Nigeria–Cameroon chimpanzees^{2,19}. Regional neighbour-joining trees and a maximum-likelihood tree estimated from allele frequencies both show that Nigeria–Cameroon and western chimpanzees form a clade. A complex demographic history has been previously reported for chimpanzees with evidence of asymmetrical gene flow among different subspecies. For instance, migration has been identified from western into eastern chimpanzees⁴, two subspecies that are currently geographically isolated. We find support for this using the D-statistic, a model-free approach that tests whether unequal levels of allele sharing between an outgroup and two populations that have more recently diverged ($D(H,W,E,C) > 16$ s.d.). However, no previous genome-wide analysis that has examined gene flow included chimpanzees from the Nigeria–Cameroon subspecies and a comparison of them with eastern chimpanzees results in a highly significant D-statistic ($D(H,E,W,N) > 25$ s.d.). Furthermore, TreeMix, a model-based approach that identifies gene flow events to explain allele frequency patterns not captured by a simple branching phylogeny, infers a signal of gene flow between Nigeria–Cameroon and

eastern chimpanzees ($P = 2 \times 10^{300}$). A more detailed treatment of gene flow applying different models and methods may be found in the Supplementary Information.

Genetic diversity is depressed at or close to genes in almost all species (Supplementary Fig. 11.1) with the effect less pronounced in subspecies with lower estimated N_e , consistent with population genetic theory. When we compare the relative level of X chromosome and autosomal (X/A) diversity across great apes as a function of genetic distance from genes, the eastern lowland gorillas and Bornean orangutans are outliers, with substantially reduced X/A diversity compared to the neutral expectation of 0.75, regardless of the distance to genes. This pattern is consistent with a recent reduction in effective population size²⁰, clearly visible in the PSMC analysis for both species (Fig. 3). However, bonobos also demonstrate a relatively constant level of X/A diversity regardless of distance from genes, with values very much in line with neutral expectations. All other subspecies demonstrate a pattern consistent with previous studies in humans²¹ where X/A diversity is lower than 0.75 close to genes and higher farther away from genes.

It has been proposed that loss of gene function may represent a common evolutionary mechanism to facilitate adaptation to changes in an environment²². There has been speculation that the success of humans may have, in part, been catalysed by an excess of beneficial

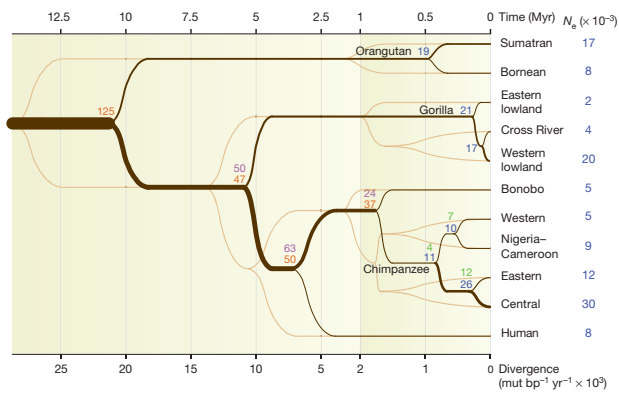


Figure 2 | Inferred population history. Population splits and effective population sizes (N_e) during great ape evolution. Split times (dark brown) and divergence times (light brown) are plotted as a function of divergence (d) on the bottom and time on top. Time is estimated using a single mutation rate (μ) of 1×10^{-9} mut bp $^{-1}$ year $^{-1}$. The ancestral and current effective population sizes are also estimated using this mutation rate. The results from several methods used to estimate N_e (COALHMM, ILS COALHMM, PSMC and ABC) are coloured in orange, purple, blue and green, respectively. The chimpanzee split times are estimated using the ABC method. The x axis is rescaled for divergences larger than 2×10^{-3} to provide more resolution in recent splits. All the values used in this figure can be found in Supplementary Table 5. The terminal N_e correspond to the effective population size after the last split event.

loss-of-function mutations²³. We thus characterized the distribution of fixed loss-of-function mutations among different species of great apes identifying nonsense and frameshift mutations resulting from SNVs ($n = 806$) and indels ($n = 1080$) in addition to gene deletion events ($n = 96$) (Supplementary Table 4). We assigned these events to the phylogeny and determined that the number of fixed loss-of-function mutations scales proportionally to the estimated branch lengths ($R^2 = 0.987$ SNVs, $R^2 = 0.998$ indels). In addition, we found no evidence of distortion on the terminal branches of the tree compared to point mutations based on a maximum likelihood analysis (Supplementary Information). Thus, the human branch in particular showed no excess of fixed loss-of-function mutations even after accounting for human-specific pseudogenes²⁴ (Supplementary Information).

Our analysis provides one of the first genome-wide views of the major patterns of evolutionary diversification among great apes. We have generated the most comprehensive catalogue of SNPs for chimpanzees (27.2 million), bonobos (9.0 million), gorillas (19.2 million) and orangutans (24.3 million) (Table 1) to date and identified several thousand AIMs, which provides a useful resource for future analyses of ape populations. Humans, western chimpanzees and eastern gorillas all show a remarkable dearth of genetic diversity when compared to other great apes. It is striking, for example, that sequencing of 79 great ape genomes identifies more than double the number of SNPs obtained from the recent sequencing of more than a thousand diverse humans²⁵—a reflection of the unique out-of-Africa origin and nested phylogeny of our species.

We provide strong genetic support for distinct populations and subpopulations of great apes with evidence of additional substructure. The common chimpanzee shows the greatest population stratification when compared to all other lineages with multiple lines of evidence supporting two major groups: the western and Nigeria–Cameroon and the central and eastern chimpanzees. The PSMC analysis indicates a temporal order to changes in ancestral effective population sizes over the last two million years, previous to which the *Pan* genus suffered a dramatic population collapse. Eastern chimpanzee populations reached their maximum size first, followed by the central and western chimpanzee. The Nigeria–Cameroon chimpanzee population size appears much more constant.

Despite their rich evolutionary history, great apes have experienced drastic declines in suitable habitat in recent years²⁶, along with declines in local population sizes of up to 75% (ref. 27). These observations highlight the urgency to sample from wild ape populations to more fully understand reservoirs of genetic diversity across the range of each species and to illuminate how basic demographic processes have affected it. The >80 million SNPs we identified in this study may now be used to characterize patterns of genetic differentiation among great apes in sanctuaries and zoos and, thus, are of great importance for the conservation of these endangered species with regard to their original range. These efforts will greatly enhance conservation planning and management of apes by providing important information on how to maintain genetic diversity in wild populations for future generations.

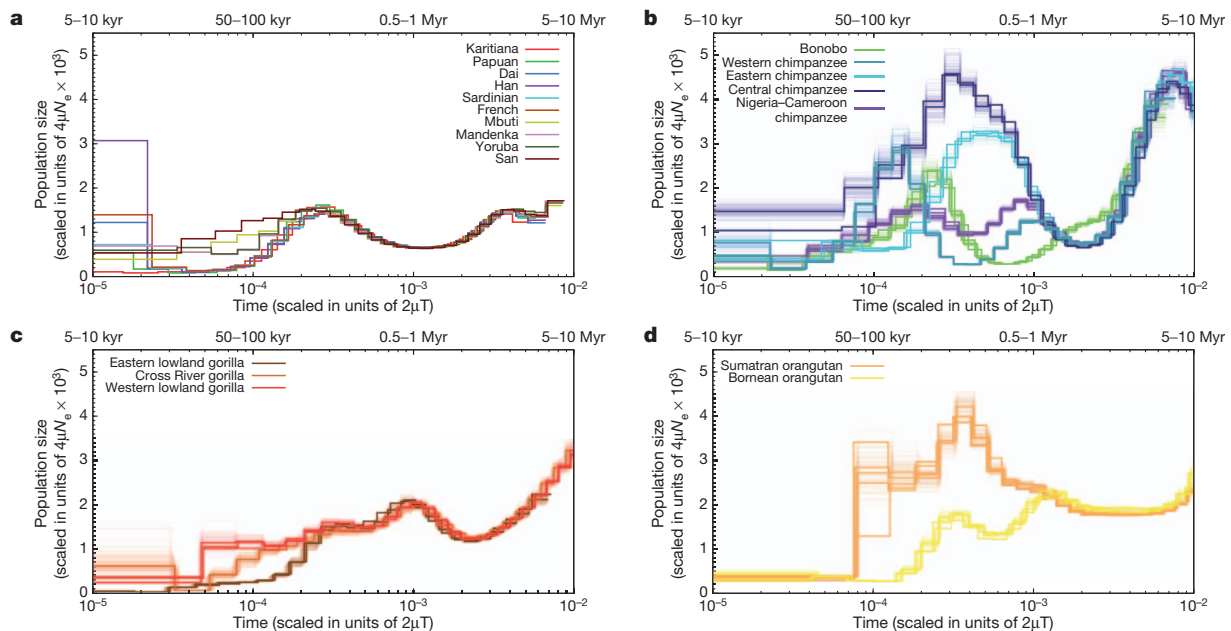


Figure 3 | PSMC analysis. Inferred historical population sizes by pairwise sequential Markovian coalescent analysis. The lower x axis gives time measured by pairwise sequence divergence and the y axis gives the effective population size measured by the scaled mutation rate. The upper x axis indicates scaling in

years, assuming a mutation rate ranging from 10^{-9} to 5×10^{-10} per site per year. The top left panel shows the inference for modern human populations. In the rest of the three panels, thin light lines of the same colour correspond to PSMC inferences on 100 rounds of bootstrapped sequences.

METHODS SUMMARY

We sequenced to a mean coverage of $25\times$ (Illumina HiSeq 2000) a total of 79 great ape individuals, representing 10 subspecies and four genera of great apes from a variety of populations across the African continent and Southeast Asia. SNPs were called using GATK¹² after BWA²⁸ mapping to the human genome (NCBI Build 36) using relaxed mapping parameters. Samples combined by species were realigned around putative indels. SNP calling was then performed on the combined individuals for each species. For indels, we used the GATK Unified Genotyper to produce an initial set of indel candidates applying several quality filters and removing variants overlapping segmental duplications and tandem repeats. We also removed groups of indels clustering within 10 bp to eliminate possible artefacts in problematic regions. Conservative allelic imbalance filters were used to eliminate false heterozygotes that may affect demographic analyses, some of which are sensitive to low levels of contamination. We estimate that the application of this filter resulted in a 14% false negative rate for heterozygotes. Our multispecies study design facilitated this assessment of contamination, which may remain undetected in studies focused on assessing diversity within a single species. The amount of cross-species contamination was estimated from the amount of non-endogenous mitochondrial sequence present in an individual. Because we wished to compare patterns of variation between and within species, we report all variants with respect to coordinates of the human genome reference. For FRAPPE analyses, we used MAF0.06 (human, orangutan and bonobo) and 0.05 (chimpanzee and gorilla) to remove singletons. For most of the analyses, we only used autosomal markers, except in the X/A analysis. To determine the amount of inbreeding, we calculated the heterozygosity genome-wide in windows of 1 megabase with 200-kilobase sliding windows. We then clustered together the neighbouring regions to account for runs of homozygosity. For the PSMC analyses, we called the consensus bases using SAMtools²⁹. Underlying raw sequence data are available through the Sequence Read Archive (SRA) (PRJNA189439 and SRP018689). Data generated in this work are available from (<http://biologiaevolutiva.org/greatape/>). A complete description of the materials and Methods is provided in the Supplementary Information.

Received 30 December 2012; accepted 26 April 2013.

Published online 3 July 2013.

- Gonder, M. K. *et al.* A new west African chimpanzee subspecies? *Nature* **388**, 337 (1997).
- Gonder, M. K. *et al.* Evidence from Cameroon reveals differences in the genetic structure and histories of chimpanzee populations. *Proc. Natl Acad. Sci. USA* **108**, 4766–4771 (2011).
- Hobolth, A., Christensen, O. F., Mailund, T. & Schierup, M. H. Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS Genet.* **3**, e7 (2007).
- Hey, J. The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Mol. Biol. Evol.* **27**, 921–933 (2010).
- Becquet, C. & Przeworski, M. A new approach to estimate parameters of speciation models with application to apes. *Genome Research* **17**, 1505–1519 (2007).
- Mailund, T. *et al.* A new isolation with migration model along complete genomes infers very different divergence processes among closely related great ape species. *PLoS Genet.* **8**, e1003125 (2012).
- Coop, G. & Przeworski, M. An evolutionary view of human recombination. *Nature Rev. Genet.* **8**, 23–34 (2007).
- Bustamante, C. D. *et al.* Natural selection on protein-coding genes in the human genome. *Nature* **437**, 1153–1157 (2005).
- Hahn, B. H. AIDS as a zoonosis: scientific and public health implications. *Science* **287**, 607–614 (2000).
- Keele, B. F. *et al.* Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* **313**, 523–526 (2006).
- Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
- DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43**, 491–498 (2011).
- Pemberton, T. J. *et al.* Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.* **91**, 275–292 (2012).
- Tang, H., Peng, J., Wang, P. & Risch, N. J. Estimation of individual admixture: analytical and study design considerations. *Genet. Epidemiol.* **28**, 289–301 (2005).
- Nsubuga, A. M., Holzman, J., Chemnick, L. G. & Ryder, O. A. The cryptic genetic structure of the North American captive gorilla population. *Conserv. Genet.* **11**, 161–172 (2010).
- Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
- Beaumont, M. A., Zhang, W. & Balding, D. J. Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025–2035 (2002).
- Morin, P. A. *et al.* Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science* **265**, 1193–1201 (1994).
- Bjork, A., Liu, W., Wertheim, J. O., Hahn, B. H. & Worobey, M. Evolutionary history of chimpanzees inferred from complete mitochondrial genomes. *Mol. Biol.* **28**, 615–623 (2011).
- Pool, J. E. & Nielsen, R. Population size changes reshape genomic patterns of diversity. *Evolution* **61**, 3001–3006 (2007).
- Hammer, M. F. *et al.* The ratio of human X chromosome to autosome diversity is positively correlated with genetic distance from genes. *Nature Genet.* **42**, 830–831 (2010).
- Olson, M. V. & Varki, A. Sequencing the chimpanzee genome: insights into human evolution and disease. *Nature Rev. Genet.* **4**, 20–28 (2003).
- Olson, M. V. When less is more: gene loss as an engine of evolutionary change. *Am. J. Hum. Genet.* **64**, 18–23 (1999).
- Wang, X., Grus, W. E. & Zhang, J. Gene losses during human origins. *PLoS Biol.* **4**, e52 (2006).
- The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
- Junker, J. *et al.* Recent decline in suitable environmental conditions for African great apes. *Divers. Distrib.* **18**, 1077–1091 (2012).
- Campbell, G., Kuehl, H., N'Goran Kouamé, P. & Boesch, C. Alarming decline of West African chimpanzees in Côte d'Ivoire. *Curr. Biol.* **18**, R903–R904 (2008).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the following funding agencies: ERC Starting Grant (260372) to T.M.-B.; NIH grants HG002385 to E.E.E., R01_HG005226 to K.R.V., A.E.W., M.F.H., L.S. and J.D.W., GM100233 and NSF HOMINID grant 1032255 to D.R. and He.Li.; MICINN (Spain) BFU2011-28549 to T.M.-B., BFU2010-19443 to Ja.Be., Spanish Government and FEDER for grants BFU2009-13409-C02-02 and BFU2012-38236 to A.N. and J.P.-M., Direcció General de Recerca, Generalitat de Catalunya (Grup de Recerca Consolidat 2009 SGR 1101) to Ja.Be., D.C., A.N. and T.M.-B.; ERC Advanced Grant (233297) and Max Planck Society to S. Paabo; Danish Council for Independent Research Natural Sciences to H.S.; Spanish Grant (CGL-2010-20170) and Zoo de Barcelona (Beca PRIC) to A.R.-H.; EUPRIM-Net to BPRC; DP1ES02577-04 NIH grant to S.A.T.; NSF Grant 0755823 to M.K.G.; P.G. is supported by the G. Harold and Leila Y. Mathers Foundation. A.N. and T.M.-B. are ICREA Research Investigators (Institut Català d'Estudis i Recerca Avancats de la Generalitat de Catalunya). J.P.-M. is supported by the Zoo de Barcelona and l'Ajuntament de Barcelona. P.H.S. is supported by an HHMI International Student Fellowship. E.E.E. is an investigator with the Howard Hughes Medical Institute. We are especially grateful to all those who generously provided the samples for the project: O. Thalmann and H. Siedel from Limbe Sanctuary; R. Garriga from Tacugama Sanctuary; W. Schempp (University of Freiburg), Burgers' Zoo; Zoo of Antwerp; Wilhelma Zoo; Givskud Zoo; Ngamba Island Chimpanzee Sanctuary and Centre de Primatologie; Centre International de Recherches Médicales de Franceville; North Carolina Zoological Park; Zoo Atlanta; the Lincoln Park Zoo (Chicago); the Antwerp Zoo and the Limbe Wildlife Centre (Cameroon); D. Travis from University of Minnesota and M. Kinsel from University of Illinois Urbana-Champaign and S. Paabo and L. Vigilant, Max Planck Institute for Evolutionary Anthropology. We thank T. Brown for revising the manuscript, L. Capilla and E. Eyraas for technical support, and M. Dierssen for comments on genes expressed in the brain.

Author Contributions E.E.E. and T.M.-B. designed the study. J.P.-M., P.H.S., J.M.K., J.L.K., B.L.-G., M.D., M.F.-C., J.C.M., C.D.B., E.E.E. and T.M.-B. analysed the raw data and performed the variant calling. J.P.-M., P.H.S., Ma.Ma., J.H.-R., I.H.-H., T.C., C.B., L.V., A.R.-H. and C.C. validated the different variants. J.P.-M., P.H.S., B.L.-G., C.A., F.H., E.E.E. and T.M.-B. analysed large deletions. K.R.V., L.J., A.E.W. and M.F.H. analysed the X/Autosome diversity. D.T., G.S., A.C., C.T., F.C., Ha.La., K.P., M.P., M.L., N.P., D.C., Ja.Be., A.N. and A.M.A. performed selection analyses. J.P.-M., P.H.S., J.M.K., J.L.K., T.D.O'C., He.Li., D.R., K.M., A.H., A.E.H., M.H.S., C.H., J.M.A., T.M., C.D.B., E.E.E. and T.M.-B. analysed different aspects of demography. M.L.W., L.S., T.A., I.K., A.P., F.L., J.A.K., E.L., P.G., H.S., M.K.G., S.A.T., R.A.B., R.E.B., O.A.R. and B.H.H. provided critical samples and participated in the discussion of phylogeny. L.F., R.K.W., Ju.Bi., E.E.E., Ma.Ma., L.A.-C., M.G. and I.G.G. generated genome libraries and produced the genome sequence associated with this project. All authors contributed to data interpretation. J.P.-M., P.H.S., E.E.E. and T.M.-B. drafted the manuscript with input from all authors.

Author Information Underlying raw sequence data are available through the Sequence Read Archive (SRA) (PRJNA189439 and SRP018689). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.M.-B. (tomas.marques@upf.edu) or E.E.E. (eee@gs.washington.edu).



This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

North Atlantic Ocean control on surface heat flux on multidecadal timescales

Sergey K. Gulev^{1,2,3}, Mojib Latif^{2,4}, Noel Keenlyside⁵, Wonsun Park² & Klaus Peter Koltermann³

Nearly 50 years ago Bjerknes¹ suggested that the character of large-scale air–sea interaction over the mid-latitude North Atlantic Ocean differs with timescales: the atmosphere was thought to drive directly most short-term—interannual—sea surface temperature (SST) variability, and the ocean to contribute significantly to long-term—multidecadal—SST and potentially atmospheric variability. Although the conjecture for short timescales is well accepted, understanding Atlantic multidecadal variability (AMV) of SST^{2,3} remains a challenge as a result of limited ocean observations. AMV is nonetheless of major socio-economic importance because it is linked to important climate phenomena such as Atlantic hurricane activity and Sahel rainfall, and it hinders the detection of anthropogenic signals in the North Atlantic sector^{4–6}. Direct evidence of the oceanic influence of AMV can only be provided by surface heat fluxes, the language of ocean–atmosphere communication. Here we provide observational evidence that in the mid-latitude North Atlantic and on timescales longer than 10 years, surface turbulent heat fluxes are indeed driven by the ocean and may force the atmosphere, whereas on shorter timescales the converse is true, thereby confirming the Bjerknes conjecture. This result, although strongest in boreal winter, is found in all seasons. Our findings suggest that the predictability of mid-latitude North Atlantic air–sea interaction could extend beyond the ocean to the climate of surrounding continents.

AMV is characterized by coherent changes in SST over the whole of the North Atlantic that are most pronounced in the extratropics with a period of 70–80 years^{3,4}. In many climate models, AMV results from variations in the Atlantic Meridional Overturning circulation (AMOC) that are generated internally by the coupled ocean–atmosphere system itself. Some studies^{7,8}, however, argue that radiative forcing by aerosols drives AMV, but the impact of aerosols on climate remains highly uncertain⁹. Although AMV has been shown to influence tropical climate⁶ and extratropical summer conditions in the North Atlantic sector⁵, the role of the mid-latitude SST variations in forcing a large-scale atmospheric response is controversial^{10,11}. Nevertheless, there is some evidence from data and climate models that AMV forces coordinated variations in European and North American climate^{5,12–14} through the advection of heat and moisture released from the North Atlantic.

Air–sea heat exchange is at the core of the whole chain of processes by which the North Atlantic Ocean may influence the atmosphere. If mid-latitude decadal and longer-term variations in SST are driven largely by ocean dynamics, then increasing or decreasing SST results in enhanced or reduced heat release, respectively, from the ocean to the atmosphere. The opposite process dominates air–sea flux variability on short timescales¹⁵: increasing or decreasing air–sea fluxes extract heat from the ocean and imply a decrease or increase, respectively, in SST. To explicitly quantify the transition between the short-term SST variations forced by the atmosphere and long-term oceanic changes that may influence the atmosphere, one therefore has to investigate surface heat fluxes. However, in contrast to SST and air temperature,

temporally homogeneous long time series of surface fluxes were not available as a result of only poorly sampled parameters needed for flux computations, especially before the 1950s¹⁶. Here we use a new data set of turbulent heat fluxes reconstructed for the period from 1880 onwards from exclusively voluntary observing ship (VOS) observations of surface meteorological variables. This long record enables us to explicitly test the Bjerknes conjecture with data.

An AMV index is computed by averaging detrended monthly SST¹⁷ anomalies over the Atlantic region 35–50°N (Fig. 1a). Although this region differs from those used for quantifying AMV previously^{3,4,6}, SST-based AMV index defined for this region is highly correlated with alternative definitions (Supplementary Fig. 1 and Supplementary Table 1) and depicts strong multidecadal variability. The major motivation for choosing this region is the larger number of observations of meteorological variables used for reconstructing surface turbulent heat fluxes. Time series of surface turbulent heat fluxes for the North Atlantic from 1880 onwards and at 5° × 5° spatial resolution are derived from VOS observations by the procedure of homogenization of sampling, a simplified bulk algorithm for the computation of fluxes¹⁸, and a priori knowledge of the statistical distribution of surface turbulent heat fluxes¹⁹ (Methods). Climatological total (sensible plus latent) turbulent heat fluxes over the North Atlantic for the period 1880–2007 capture the well-known major features, specifically the maxima over the Gulf Stream and in the subpolar North Atlantic (Fig. 1a).

The AMV index and surface turbulent heat fluxes averaged over 5° boxes in the mid-latitude and subpolar North Atlantic were decomposed into long-term (multidecadal) and short-term (interannual to decadal) components (Methods). Multidecadal fluctuations in surface turbulent heat fluxes and the SST index (as given by 11-year running means) are positively correlated (Figs 1b and 2) over the western mid-latitude North Atlantic, with the largest correlations of 0.77 found southeast of Newfoundland, which is statistically significant at the 95% level (Methods). Thus, in this region and on multidecadal timescales, increasing or decreasing surface ocean temperature goes along with a respective increase or decrease in surface turbulent fluxes; that is, the ocean heats or cools the atmosphere, respectively. This pattern holds for both sensible and latent flux (Supplementary Fig. 2) and is very persistent throughout the year, with only minor variations in the correlation strength (Supplementary Fig. 3). Similar analyses performed with NOAA-CIRES 20th Century Reanalysis V2 (20CRv2)²⁰ confirm this finding (Supplementary Fig. 4). The close association between large-scale SST anomalies and the local surface heat flux in this area is also consistent with modelling studies^{2,21}.

According to previous diagnostic and modelling work^{15,22} and as outlined above, we note that, on interannual to decadal timescales, the interaction between the SST and surface heat flux is different from that on multidecadal timescales. In mid-latitudes, the atmosphere, as a fast dynamical system, determines short-period synoptic variations in the heat flux that drive the temperature of the upper ocean mixed layer. Furthermore, the intensity of the day-to-day (synoptic) activity in the

¹P. P. Shirshov Institute of Oceanology, 36 Nakhimovsky prospect, 117997 Moscow, Russia. ²GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany.

³Moscow State University, Faculty of Geography, Leninskie Gory, 119991 Moscow, Russia. ⁴Kiel University, Cluster of Excellence 'The Future Ocean', Christian-Albrechts-Platz 4, 24118 Kiel, Germany.

⁵Geophysical Institute and Bjerknes Centre for Climate Research, University of Bergen, Allégaten 70, 5007 Bergen, Norway.

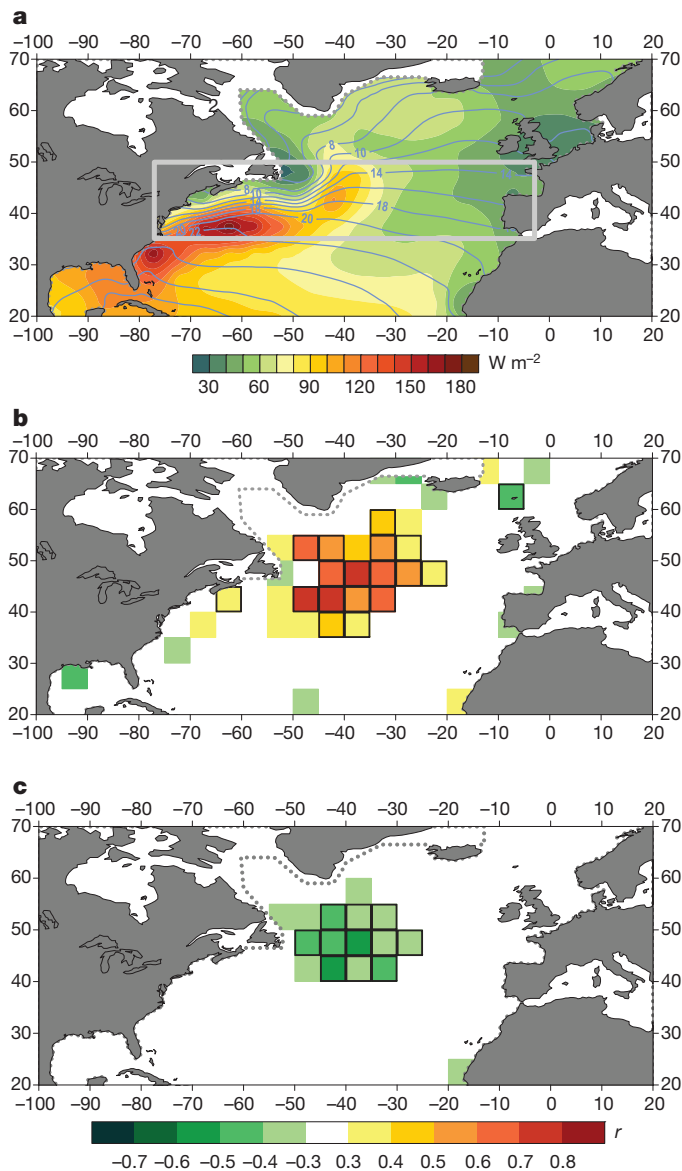


Figure 1 | Spatial pattern of correlation between the AMV SST index and anomalies of surface turbulent heat fluxes for the long-term and short-term components. **a**, SST (blue contours) and surface turbulent sensible plus latent heat flux (shading) climatologies for the period 1880–2007. Surface heat flux climatology is based on $5^\circ \times 5^\circ$ averaged time series. **b**, **c**, Correlation between the AMV SST index and local anomalies of surface turbulent sensible plus latent heat flux for the interdecadal long-term component (**b**) and the interannual short-term component (**c**). The AMV SST index is computed as SST anomalies averaged over the region $35\text{--}50^\circ\text{N}$ (grey box in **a**). For correlations in **b** and **c**, black boxes indicate 95% significance level according to a random phase test²⁸.

North Atlantic mid-latitudes is closely linked to the North Atlantic Oscillation (NAO)²³, the leading mode of internal atmospheric variability in this region, which is strongly correlated with surface fluxes on short interannual to intra-decadal timescales²⁴.

Consistent with this notion, deviations of the AMV index and surface turbulent heat flux from their respective 11-year running means, representing the short-term variability, are negatively correlated over the North Atlantic (Fig. 1c). The strongest correlations, close to -0.5 , are identified east of Newfoundland and in the southern Labrador Sea. This implies that stronger surface fluxes result in decreasing SST, whereas weaker fluxes result in increasing SST, which is especially pronounced during the cold season (Supplementary Fig. 3) when atmospheric synoptic variability is most intense. Short-term variations in North Atlantic

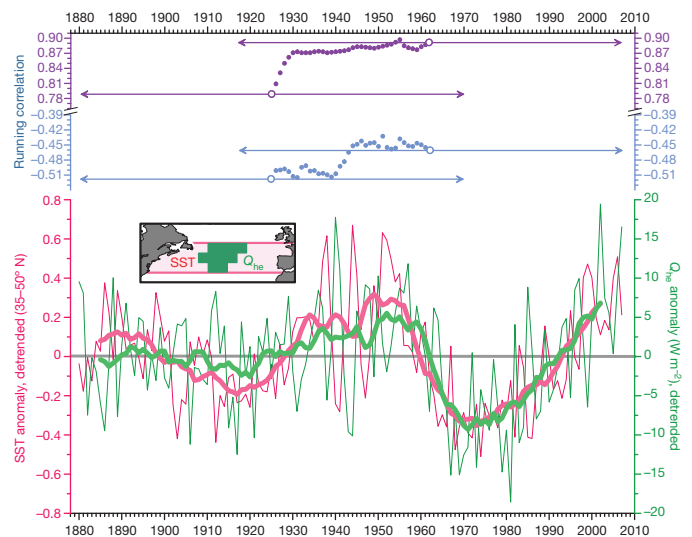


Figure 2 | Time series of AMV SST index and anomalies of sensible plus latent heat flux in the mid-latitudinal North Atlantic. Bottom: a 128-year time series of the AMV SST index derived from HadISST1 data set for $35\text{--}50^\circ\text{N}$ (red) and anomalies of surface turbulent sensible plus latent heat flux Q_{he} (green) averaged over the region shown in green in the inset map. Thin and thick lines are annual means and 11-year running means, respectively. Top: running correlation between the time series for the long-term (magenta) and short-term (blue) components. Correlations were computed over 90-year segments (indicated by horizontal arrows) moving from the beginning to the end of the records.

SST and turbulent heat flux averaged over $35\text{--}50^\circ\text{N}$ are in antiphase, especially during the periods from the 1930s to the early 1950s and after 1960 (Fig. 2). Our results for both long-term and short-term variability remain practically unchanged when the SST-based AMV index is derived by using alternative SST data or by removing the anthropogenic signal using regression of the local SST onto time series of global mean surface temperature (Supplementary Figs 5 and 6). The correlation for the long-term component is somewhat stronger during the second part of the record, whereas the anticorrelation of short-term variability is slightly stronger before 1970 (Fig. 2 and Supplementary Fig. 7). For the long-term component this is primarily due to the impact of the very beginning of the record before 1885 and can be also partly explained by the larger magnitude of interdecadal variation in both SST and surface fluxes during the last several decades (Fig. 2) and a slight displacement of the maximum correlation pattern eastward (Supplementary Fig. 7). The use of AMV indices based on alternative approaches for processing the SST data, including methods to remove anthropogenic signal, results in smaller differences in correlation between the first and second parts of the record (Supplementary Fig. 8).

The different character of air–sea interaction on multidecadal and interannual timescales is clearly demonstrated by cross-spectral analysis (Fig. 3). The spectra of North Atlantic SST and surface turbulent heat fluxes (Methods) both show a peak at a period of 50–70 years. Both multidecadal and interannual variations of SST and observationally based surface fluxes are highly coherent (Fig. 3). However, the phase lag between the two is very close to zero in the band 30–70 years and sharply switches to antiphase, with a phase of about 180° at periods smaller than decadal. Thus, there is a negative feedback such that multidecadal SST variations are damped by the surface heat fluxes, which in turn implies that the ocean heats or cools the atmosphere on these long timescales, a situation that exists on interannual timescales in parts of the Tropics, for example, the equatorial Pacific Ocean.

The timescale on which the character of air–sea interaction in the North Atlantic mid-latitudes changes qualitatively is clearly identified by the correlations between SST and surface heat fluxes calculated at different timescales and averaged over the mid-latitudinal North Atlantic (Fig. 4). The long-term components are highly positively

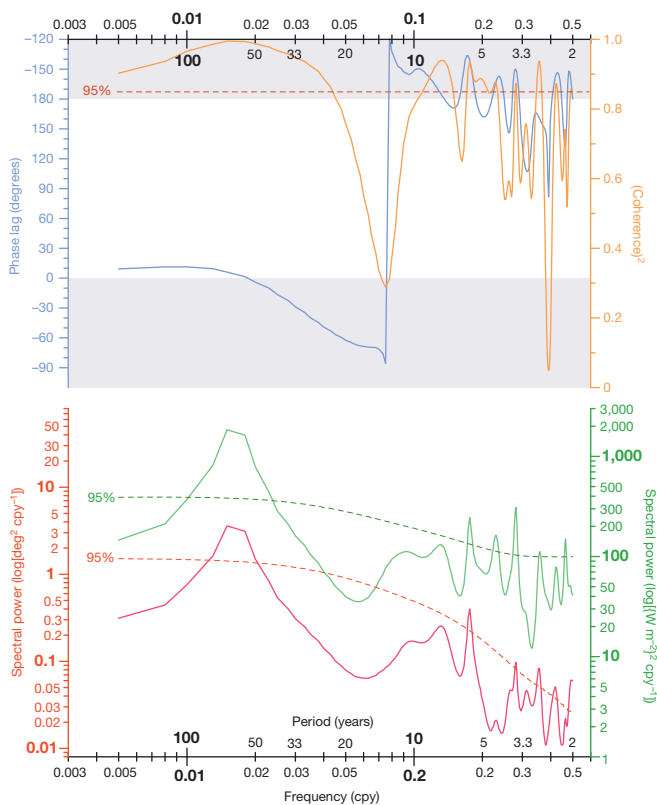


Figure 3 | Cross-spectral analysis of the AMV SST index and anomalies of surface turbulent heat fluxes. Bottom: power spectra of the SST index (red) and surface turbulent sensible plus latent heat flux in the region shown in Fig. 2 (green). There is a pronounced peak at a roughly 60-yr period in both spectra. Top: squared coherence of SST AMV index and surface heat flux (orange) and phase lag between the time series (blue). A phase lag close to zero indicates that oscillations of SST and surface fluxes are in-phase, while a phase lag varying around 180° shows that the oscillation at a given frequency is in antiphase. For easier interpretation, ordinate labels change sign at 180° . Grey shading shows the range of negative phase lags. Thin dashed lines in both panels indicate 95% significance levels for spectral power and coherence. Frequencies on the x axis are in cycles per year (cpy); additionally, corresponding periods in years are indicated.

correlated and the short-term components show significant negative correlations (and field significance in this belt) when the length of the running mean window used to separate short-term and long-term components exceeds 10 years. When the length of the running mean is shorter than 10 years, the correlations for both long-term and short-term components weaken substantially, with the pattern lacking field significance (Methods).

Our analysis based on reconstructed time series of turbulent heat flux at the sea surface during 1880–2007 clearly shows that multidecadal variations in SST in the Gulf Stream extension region have an active role in air–sea interaction and thus in diabatic heating of the lower atmosphere. The link is strongest at periods of 50–70 years and is significantly associated with AMV, which is consistent with results from climate models in the absence of external forcing. On interannual timescales, however, the atmosphere, as expected, drives changes in SST, with cooling or warming of the surface ocean under the influence of intensified or weakened surface fluxes, respectively. The results are important in the context of climate predictability in the sense that the North Atlantic multidecadal SST changes may not only be predictable but may also drive a large-scale atmospheric response. This is not so for the shorter interannual and intra-decadal timescales, on which chaotic atmospheric variability drives SST.

We have considered here only the turbulent heat fluxes, which are known to contribute the largest share to the total (net) heat flux variability at the ocean's surface. In future work it would be important

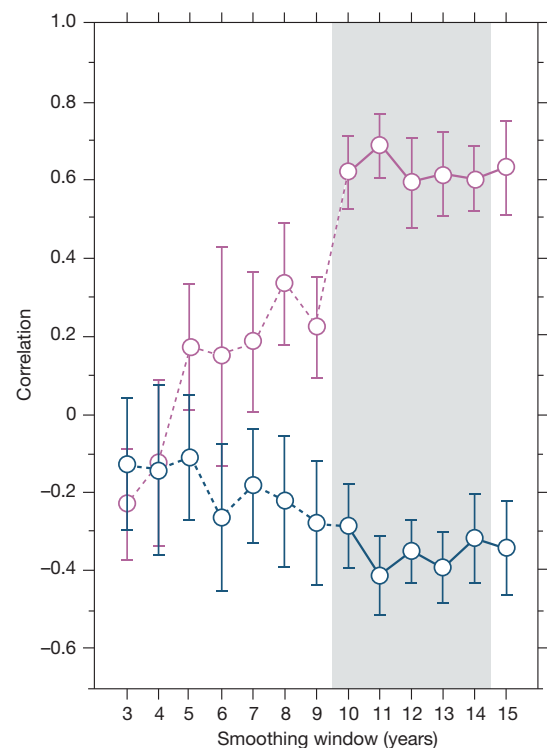


Figure 4 | Changing correlations between the AMV SST index and anomalies of surface heat fluxes with the length of the filtering window. Correlation between anomalies of the AMV SST index and surface turbulent heat fluxes were estimated with different running mean lengths for the long-term component (magenta) and the short-term component (blue). Correlation coefficients were averaged over the area shown in Fig. 2; error bars indicate s.d. of coefficient values. Highly positive correlations between the long-term components and significantly negative correlations between the short-term components are persistent for the windows of 10 years and more, implying the timescale on which the interaction between SST and surface turbulent fluxes changes its character. The grey shaded band shows the range for which patterns of positive correlations for the long-term components and negative correlations for the short-term components hold field significance within the 35–50° N area.

to extend the analysis to the radiative components of air–sea heat exchange (short-wave and long-wave radiation fluxes) to complete the surface heat balance. This will particularly permit testing of the capabilities of external radiative forcing^{7,8} as a driving factor for AMV. However, reconstruction of the radiative fluxes requires estimates of cloud cover, which was poorly sampled before World War II. Our work nevertheless opens an interesting avenue for the analysis of the results of model simulations of the climate for the previous century and will help to design new experiments, extending earlier efforts^{3,14} and targeting the role of ocean signals in forcing atmospheric variability on multidecadal timescales in the North Atlantic and possibly also other ocean basins.

METHODS SUMMARY

We have reconstructed long-term time series of surface turbulent heat fluxes using individual VOS reports from the International Comprehensive Ocean–Atmosphere Data Set (ICOADS, version 2.5)¹⁶ for the period 1880–2007. To minimize the impact of time-dependent sampling errors (Supplementary Figs 9 and 10) on the results, we subsampled^{19,25} the meteorological variables randomly, selecting 45 reports for each 5° grid cell in each season. Special selection procedures were used to minimize the impact of changes in observational practices on the results^{16,26,27}. Atmospheric humidity was reconstructed with a simple multivariate regression procedure based on the data for the period 1960–1980 (Supplementary Figs 11 and 12). Subsampling was repeated 20 times for each 5° grid cell and for each season (January–March, April–June, July–September and October–December). The robustness of the results with respect to the number of samples is discussed in Supplementary Fig. 10b, c. Surface fluxes were then computed from the subsampled data by using a simplified

version of the COARE-3.0 algorithm¹⁸ (20 subsamples). For each subsample the modified Fisher–Tippett distribution¹⁹ was used to approximate the computed fluxes and was further integrated to obtain estimates of seasonal mean flux. Averaging over 20 random subsamples provided seasonal estimates of surface sensible and latent fluxes for a grid cell.

Decomposition of the 128-year time series of the detrended (Supplementary Figs 5 and 6) SST and surface heat fluxes into long-term (interdecadal) and short-term (interannual) components was performed by applying an 11-year running mean (or running mean over different smoothing windows; see Fig. 4 for example). The correlation analysis²⁸ was performed for the detrended time series. Cross-spectral analysis was performed with the maximum entropy method²⁹ on the basis of predicting the available data to infinite time by an autoregressive fitting (Supplementary Fig. 13).

Full Methods and any associated references are available in the online version of the paper.

Received 17 March; accepted 3 May 2013.

1. Bjerknes, J. in *Advances in Geophysics* Vol. 10 (eds Landberg, H. E. & van Mieghem, J.) 1–82 (Academic, 1964).
2. Delworth, T. S., Manabe, S. & Stouffer, R. J. Interdecadal variations of the thermohaline circulation in a coupled ocean–atmosphere model. *J. Clim.* **6**, 1993–2011 (1993).
3. Latif, M. *et al.* Reconstructing, monitoring, and predicting multidecadal-scale changes in the North Atlantic thermohaline circulation with sea surface temperature. *J. Clim.* **17**, 1605–1614 (2004).
4. Knight, J. R., Allan, R. J., Folland, C. F., Vellinga, M. & Mann, M. E. A signature of persistent natural thermohaline circulation cycles in observed climate. *Geophys. Res. Lett.* **32**, L20708, <http://dx.doi.org/10.1029/2005GL024233> (2005).
5. Sutton, R. T. & Hodson, D. Climate Atlantic Ocean forcing of North American and European summer. *Science* **309**, 115–118 (2005).
6. Zhang, R. & Delworth, T. L. Impact of Atlantic multidecadal oscillations on India/Sahel rainfall and Atlantic hurricanes. *Geophys. Res. Lett.* **33**, L17712, <http://dx.doi.org/10.1029/2006GL026267> (2006).
7. Otterå, O. H., Bentsen, M., Drange, H. & Suo, L. External forcing as a metronome for Atlantic multidecadal variability. *Nature Geosci.* **3**, 688–694 (2010).
8. Booth, B. B. B., Dunstone, N. J., Halloran, P. R., Andrews, T. & Bellouin, N. Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature* **484**, 228–232 (2012).
9. Forster, P. V. *et al.* in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon, S. *et al.*) 129–234 (Cambridge Univ. Press, 2007).
10. Kushnir, Y. *et al.* Atmospheric GCM response to extratropical SST anomalies: synthesis and evaluation. *J. Clim.* **15**, 2233–2256 (2002).
11. Rodwell, M. J., Rowell, D. P. & Folland, C. K. Oceanic forcing of the wintertime North Atlantic Oscillation and European climate. *Nature* **398**, 320–323 (1999).
12. Kushnir, Y. Interdecadal variations in the North Atlantic sea surface temperature and associated atmospheric conditions. *J. Clim.* **7**, 141–157 (1994).
13. Semenov, V. A. *et al.* The impact of North Atlantic–Arctic multidecadal variability on Northern Hemisphere surface air temperature. *J. Clim.* **23**, 5668–5677 (2010).
14. Pohlmann, H., Sienz, F. & Latif, M. Influence of the multidecadal Atlantic meridional overturning circulation variability on European climate. *J. Clim.* **19**, 6062–6067 (2006).
15. Cayan, D. R. Latent and sensible heat flux anomalies over the northern oceans: driving the sea surface temperature. *J. Phys. Oceanogr.* **22**, 859–881 (1992).
16. Woodruff, S. D. *et al.* ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive. *Int. J. Climatol.* **31**, 951–967 (2011).
17. Rayner, N. A. *et al.* Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**, (D14), 4407, <http://dx.doi.org/10.1029/2002JD002670> (2003).
18. Fairall, C. W., Bradley, E. F., Hare, J. E., Grachev, A. A. & Edson, J. B. Bulk parameterization of air–sea fluxes: updates and verification for the COARE algorithm. *J. Clim.* **16**, 591 (2003).
19. Gulev, S. K. & Belyaev, K. P. Probability distribution characteristics for surface air–sea turbulent heat fluxes over the global ocean. *J. Clim.* **25**, 184–206 (2012).
20. Compo, G. P. *et al.* The Twentieth Century Reanalysis Project. *Q. J. R. Meteorol. Soc.* **137**, 1–28 (2011).
21. Shaffrey, L. & Sutton, R. Bjerknes compensation and the decadal variability of the energy transports in a coupled climate model. *J. Clim.* **19**, 1167–1181 (2006).
22. Yu, B., Boer, G. J., Zwiers, F. W. & Merryfield, W. J. Covariability of SST and surface heat fluxes in reanalyses and CMIP3 climate models. *Clim. Dyn.* **36**, 589–605 (2011).
23. Hurrell, J. W. Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science* **269**, 676–679 (1995).
24. Eden, C. & Willebrand, J. Mechanism of interannual to decadal variability of the North Atlantic circulation. *J. Clim.* **14**, 2266–2280 (2001).
25. Gulev, S. K., Jung, T. & Ruprecht, E. Estimation of the impact of sampling errors in the VOS observations on air–sea fluxes. Part I. Uncertainties in climate means. *J. Clim.* **20**, 279–301 (2007).
26. Kent, E. C., Woodruff, S. D. & Berry, D. I. Metadata from WMO Publication No. 47 and an assessment of Voluntary Observing Ship observation heights in ICOADS. *J. Atmos. Ocean. Technol.* **24**, 214–234 (2007).
27. Josey, S., Kent, E. C. & Taylor, P. K. New insights into the ocean heat budget closure problem from analysis of the SOC air–sea flux climatology. *J. Clim.* **12**, 2856–2880 (1999).
28. Ebisuzaki, W. A method to estimate the statistical significance of a correlation when the data are serially correlated. *J. Clim.* **10**, 2147–2153 (1997).
29. Hayashi, Y. Space time cross spectral analysis using the maximum entropy method. *J. Meteorol. Soc. Jpn.* **59**, 621–624 (1981).

Supplementary Information is available in the online version of the paper.

Acknowledgements This study was supported by the Deutsche Forschungsgemeinschaft under grant KE 1471/2-1 and by the Russian Ministry of Education and Science through the Special Grant for establishing excellence at Russian Universities, no. 11.G34.31.0007. We also benefited from contracts 2011-16-420-1-001 and 11.519.11.6034 with the Russian Ministry of Education and Science and the RACE project of the German Federal Ministry of Education and Research.

Author Contributions The main idea of this study belongs to S.K.G. and M.L.; most of the computations were performed by S.K.G., S.K.G., M.L., N.K., W.P. and K.P.K. have contributed equally to discussion of the results and writing of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.K.G. (gul@sail.msk.ru).

METHODS

Reconstruction of surface turbulent heat flux time series. To reconstruct long-term time series of surface turbulent heat fluxes in the North Atlantic we applied a methodology designed to minimize temporal inhomogeneities in sampling density and observational practices. We used individual VOS reports from the International Comprehensive Ocean–Atmosphere Data Set (ICOADS, version 2.5)¹⁶ for the period 1880–2007. Although North Atlantic mid-latitudes are generally better sampled with VOS observations than other regions of the World Ocean, there is a strong inhomogeneity in the number of samples through the years with the number of observations in the 1880s–1900s being 10–20-fold smaller than in the 1960s–2000s (Supplementary Figs 9 and 10). To minimize the impact of time-dependent sampling error on the results we subsampled^{19,25} meteorological variables randomly, selecting 45 reports for each 5° grid cell in each calendar season (January–March, April–June, July–September and October–December). To minimize the impact of changes in observational practices^{26,27} on the results we selected only reports with bucket measurements of SST¹⁷ and with Beaufort estimates of wind speed that were converted to wind speed by using the Lindau equivalent scale²⁷. If during the previous decades there were not enough reports meeting these requirements, information on anemometer heights of marine carriers²⁷ was used for the adjustment of wind speed to a standard height, and neutral stability and non-bucket SST measurements were corrected²⁷. Atmospheric humidity data are more poorly sampled than the other variables¹⁶, and temporal sampling inhomogeneity for humidity is much stronger than for the other parameters. To ensure homogeneity of surface turbulent heat flux estimates, instead of using actual humidity data we developed a simple multivariate regression procedure for the reconstruction of relative humidity using temperature, sea level pressure and wind speed data. The procedure was developed with the data for the period 1960–1980 (the best-sampled period for the North Atlantic), and the derived regression model was applied for the reconstruction of relative humidity for the whole period from 1880 to 2007. Use of the reconstructed humidity may result in a random error in latent heat flux estimates of about $\pm 6 \text{ W m}^{-2}$ for monthly values (Supplementary Figs 11 and 12). Reconstruction of the humidity resulted in a close correlation of the latent and sensible heat fluxes, so that we show here the results for the total (sensible plus latent) turbulent heat flux. Subsampled meteorological variables were used to compute sensible and latent heat flux estimates by using a simplified version of the COARE-3.0 algorithm¹⁸, with parts of the scheme requiring additional parameters being neglected. Subsampling procedures were repeated 20 times for each 5° grid cell and for each calendar season. For every subsample a statistical distribution of the computed fluxes was then approximated by the modified Fisher–Tippett distribution¹⁹, which was integrated to obtain seasonal mean flux estimates. The average over 20 simulations of the random subsampling provided seasonal estimates of surface sensible and latent heat fluxes for a grid cell. We believe that this procedure produces North Atlantic surface turbulent heat flux time series that are homogeneous in terms of sampling and observational practices. To investigate the impact of the number of samples on the results, we repeated calculations for different numbers of samples from 20 to 70 per 5° grid cell per season. Supplementary Fig. 10b, c shows the robustness of the

results to the number of samples. For some regions in the late nineteenth century and the first half of the twentieth century the number of observations was nevertheless very small for deriving the values of surface heat fluxes. For these locations heat flux anomalies were produced by an interpolation procedure (Supplementary Fig. 14 and caption to Supplementary Fig. 5). Comparison of the reconstructed turbulent heat fluxes with the other advanced surface heat flux data sets³⁰ for the past few decades shows that our fluxes are generally smaller by about 10–30%, although their interannual variability is quite comparable (Supplementary Fig. 15).

Statistical methods. Decomposition of 128-year time series of SST and surface fluxes into long-term (interdecadal) and short-term (interannual) components was performed by applying an 11-year running mean (or a running mean over different smoothing windows, for example Fig. 4). The running mean was considered to be the long-term component, and deviations from it were considered to be the short-term component. More sophisticated filtering procedures based on different low-pass and high-pass filters including Lanczos filtering³¹ were also applied, with very similar results. However, these procedures result typically in a much larger time series cutoff—that is, data loss—of time series compared with the running mean. Thus, all analyses shown here were performed with running-mean filtering.

Because we focused here on interdecadal and interannual variability and not on secular signals, the time series of both SST and surface turbulent fluxes were detrended by removing the linear trends, estimated with the least-squares procedure. An alternative method to account for secular changes potentially driven by externally forced climate change was to compute an AMV SST index from SST by first subtracting (spatially varying) SST variations that were linearly related to global mean surface temperature³². Although differences exist between the two indices, our key findings remain unchanged (Supplementary Figs 6 and 8). The correlation analysis was performed for the detrended time series. The significance of correlation for long-term and short-term components was estimated by taking into account autocorrelation of the time series with the non-parametric test based on phase-scrambling bootstrapping in the frequency domain²⁸.

Cross-spectral analysis was performed with the maximum-entropy method²⁹ based on predicting the available data to infinite time by an autoregressive fitting. An advantage of this method is that it is applicable to short time series when oscillations with two or three periods per record should be identified. The results of the spectral analysis are practically insensitive to the order of the autoregressive model²⁹, implying the robustness of estimates of spectral peaks, coherence and phase (Supplementary Fig. 13). Estimates of field significance of correlation patterns were performed from the binomial distribution³³ for the area of the mid-latitudinal North Atlantic from 35° N to 50° N.

30. Yu, L. & Weller, R. A. Objectively analyzed air–sea heat fluxes for the global ice-free oceans (1981–2005). *Bull. Am. Meteorol. Soc.* **88**, 527–539 (2007).
31. Duchon, C. E. Lanczos filtering in one and two dimensions. *J. Appl. Meteor.* **18**, 1016–1022 (1979).
32. Ting, M., Kushnir, Y., Seager, R. & Li, C. Forced and internal twentieth-century SST trends in the North Atlantic. *J. Clim.* **22**, 1469–1481 (2009).
33. Livezey, R. E. & Chen, W. Y. Statistical field significance and its determination by Monte-Carlo techniques. *Mon. Weath. Rev.* **111**, 46–59 (1983).

Vascularized and functional human liver from an iPSC-derived organ bud transplant

Takanori Takebe^{1,2}, Keisuke Sekine¹, Masahiro Enomura¹, Hiroyuki Koike¹, Masaki Kimura¹, Takunori Ogaeri¹, Ran-Ran Zhang¹, Yasuharu Ueno¹, Yun-Wen Zheng¹, Naoto Koike^{1,3}, Shinsuke Aoyama⁴, Yasuhisa Adachi⁴ & Hideki Taniguchi^{1,2}

A critical shortage of donor organs for treating end-stage organ failure highlights the urgent need for generating organs from human induced pluripotent stem cells (iPSCs)¹. Despite many reports describing functional cell differentiation^{2–4}, no studies have succeeded in generating a three-dimensional vascularized organ such as liver. Here we show the generation of vascularized and functional human liver from human iPSCs by transplantation of liver buds created *in vitro* (iPSC-LBs). Specified hepatic cells (immature endodermal cells destined to track the hepatic cell fate) self-organized into three-dimensional iPSC-LBs by recapitulating organogenetic interactions between endothelial and mesenchymal cells⁵. Immunostaining and gene-expression analyses revealed a resemblance between *in vitro* grown iPSC-LBs and *in vivo* liver buds. Human vasculatures in iPSC-LB transplants became functional by connecting to the host vessels within 48 hours. The formation of functional vasculatures stimulated the maturation of iPSC-LBs into tissue resembling the adult liver. Highly metabolic iPSC-derived tissue performed liver-specific functions such as protein production and human-specific drug metabolism without recipient liver replacement⁶. Furthermore, mesenteric transplantation of iPSC-LBs rescued the drug-induced lethal liver failure model. To our knowledge, this is the first report demonstrating the generation of a functional human organ from pluripotent stem cells. Although efforts must ensue to translate these techniques to treatments for patients, this proof-of-concept demonstration of organ-bud transplantation provides a promising new approach to study regenerative medicine.

Since the discovery of embryonic stem cells in 1981, decades of laboratory studies have failed to generate a complex vascularized organ such as liver from pluripotent stem cells, giving rise to the prevailing belief that *in vitro* recapitulation of the complex interactions among cells and tissues during organogenesis is essentially impractical⁷. Here we challenge this idea by focusing on the earliest process of organogenesis, that is, cellular interactions during organ-bud development.

During early liver organogenesis, newly specified hepatic cells delaminate from the foregut endodermal sheet and form a liver bud⁸, a condensed tissue mass that is soon vascularized. Such large-scale morphogenetic changes depend on the exquisite orchestration of signals between endodermal epithelial, mesenchymal and endothelial progenitors before blood perfusion⁵. These observations led us to propose that three-dimensional liver-bud formation can be recapitulated *in vitro* by culturing hepatic endoderm cells with endothelial and mesenchymal lineages (Fig. 1a). To examine this hypothesis, we first prepared hepatic endoderm cells from human iPSCs (iPSC-HEs) by directed differentiation, producing approximately 80% of the treated cells expressed the hepatic marker HNF4A, which is involved in cell fate determination (Supplementary Fig. 1).

Next, to recapitulate early organogenesis, human iPSC-HEs were cultivated with stromal cell populations; human umbilical vein endothelial cells (HUVECs) and human mesenchymal stem cells (MSCs) unless stated otherwise, because of their primitive nature (Supplementary Fig. 2a and Supplementary Discussion). Notably, although cells were plated in

two-dimensional conditions, human iPSC-HEs self-organized into macroscopically visible three-dimensional cell clusters by an intrinsic organizing capacity up to 48 h after seeding (Fig. 1b, c, Supplementary

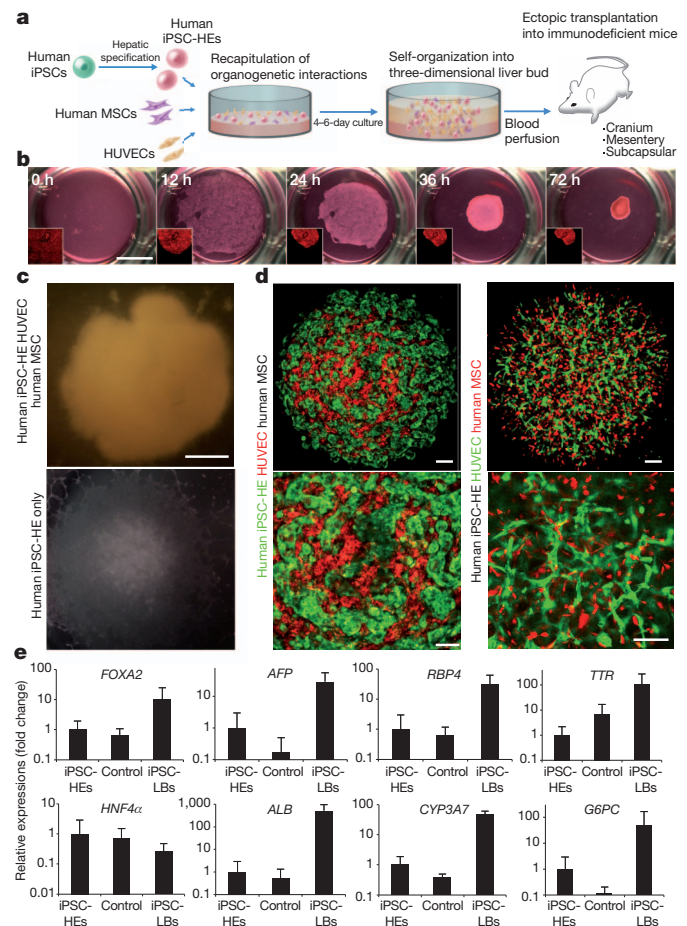


Figure 1 | Generation of human liver buds from human iPSCs. **a**, Schematic representation of our strategy. **b**, Self-organization of three-dimensional human iPSC-LBs in co-cultures of human iPSC-HEs with HUVECs and human MSCs (see also Supplementary Video 1). The time-lapse fluorescence imaging of human iPSC-HEs is shown here. Scale bar, 5 mm. **c**, Gross observation of human iPSC-LBs (top panel) and conventional two-dimensional cultures (bottom panel). Scale bar, 1 mm. **d**, Presence of human iPSC-HEs and nascent endothelial networks inside human iPSC-LBs. Green, human iPSC-HE or HUVEC; red, HUVEC or human MSC. Scale bars, 100 μ m. **e**, Quantitative PCR analysis of hepatic marker gene expression in human iPSC-LBs at day 6 of culture. Control samples were iPSC-HE:HUVEC: humanMSC cells that had been grown as separate cell types and not cultured together, then for control gene expression analysis the cells were mixed together at the same ratio of cell types as in human iPSC-LBs. Results represent mean \pm s.d., $n = 8$.

¹Department of Regenerative Medicine, Yokohama City University Graduate School of Medicine, 3-9 Fukuura, Kanazawa-ku, Yokohama, Kanagawa 236-0004, Japan. ²Advanced Medical Research Center, Yokohama City University, Yokohama, Kanagawa 236-0004, Japan. ³Department of Surgery, Seirei Sakura Citizen Hospital, 2-36-2 Ebaradai, Sakura, Chiba 285-8765, Japan. ⁴ADME & Tox. Research Institute, Sekisui Medical Company Ltd., Tokai, Ibaraki 319-1182, Japan.

Fig. 2b, c and Supplementary Video 1). The presumed human iPSC-derived liver buds (iPSC-LBs) were mechanically stable and could be manipulated physically. We visualized a formation of endothelial network and homogeneously distributed human iPSC-HEs by using fluorescent-protein-labelled cells (Fig. 1d). Although human iPSC-LBs are a tri-lineage mixed tissue and difficult to compare directly with human iPSC-HEs, quantitative polymerase chain reaction (PCR) analysis revealed that cells in human iPSC-LB had significantly increased expression of early hepatic marker genes such as alpha-fetoprotein (AFP), retinol binding protein 4 (RBP4), transthyretin (TTR) and albumin (ALB) (Fig. 1e)⁴. Microarray profiling showed that FGF and BMP pathways were upregulated highly in human iPSC-HEs when co-cultured with stromal cells, suggesting the involvement of stromal-cell-dependent factors in liver-bud formation (Supplementary Fig. 3a). Consistent with this, loss- and gain-of-function experiments suggested that, in addition to the direct cell-to-cell interactions, stromal-cell-dependent paracrine support is essential for three-dimensional liver-bud formation through the activating FGF and BMP pathways (Supplementary Fig. 3b–i and Supplementary Discussion)⁹.

Human liver-bud formation is initiated on the third or fourth week of gestation, and this corresponds to embryonic day 9.5 (E9.5) to E10.5 for mouse liver-bud formation¹⁰. Similar to E10.5 mouse liver bud, immunohistochemistry showed that human iPSC-LB is composed of proliferating AFP-positive hepatoblasts¹¹ as well as mesenchymal and endothelial

progenitors (Fig. 2a). Hepatic cells in human iPSC-LBs were as proliferative as E10.5 mouse liver buds (Fig. 2b). Flow cytometric characterization revealed that $42.7 \pm 7.5\%$ ($n = 6$) of cells in human iPSC-LBs were identified as iPSC-HEs using fluorescence-labelled HUVECs and MSCs. Among these, approximately $71.9 \pm 7.3\%$ ($n = 6$) of human iPSC-HEs expressed ALB and $29.9 \pm 2.8\%$ expressed AFP, and $19.3 \pm 4.5\%$ were positive for both ALB and AFP (Fig. 2c).

To characterize the expression profiles of human iPSC-LBs and to compare with those of a corresponding developmental stage, we carried out microarray analysis of 83 selected genes that are serially upregulated during liver development. Hierarchical clustering analyses suggested that the expression profiles of human iPSC-LBs at day 4 of culture resembled those of mouse E10.5 and E11.5 liver buds rather than advanced fetal or adult livers (Fig. 2d). These expression profiles were relatively similar to those of human fetal liver cell-derived liver buds (FLC-LBs) (Fig. 2d), which also have an ability to form LBs (Supplementary Figs 2a and 4). The 83-gene expression profile of human iPSC-LBs showed closer signatures to more advanced human liver

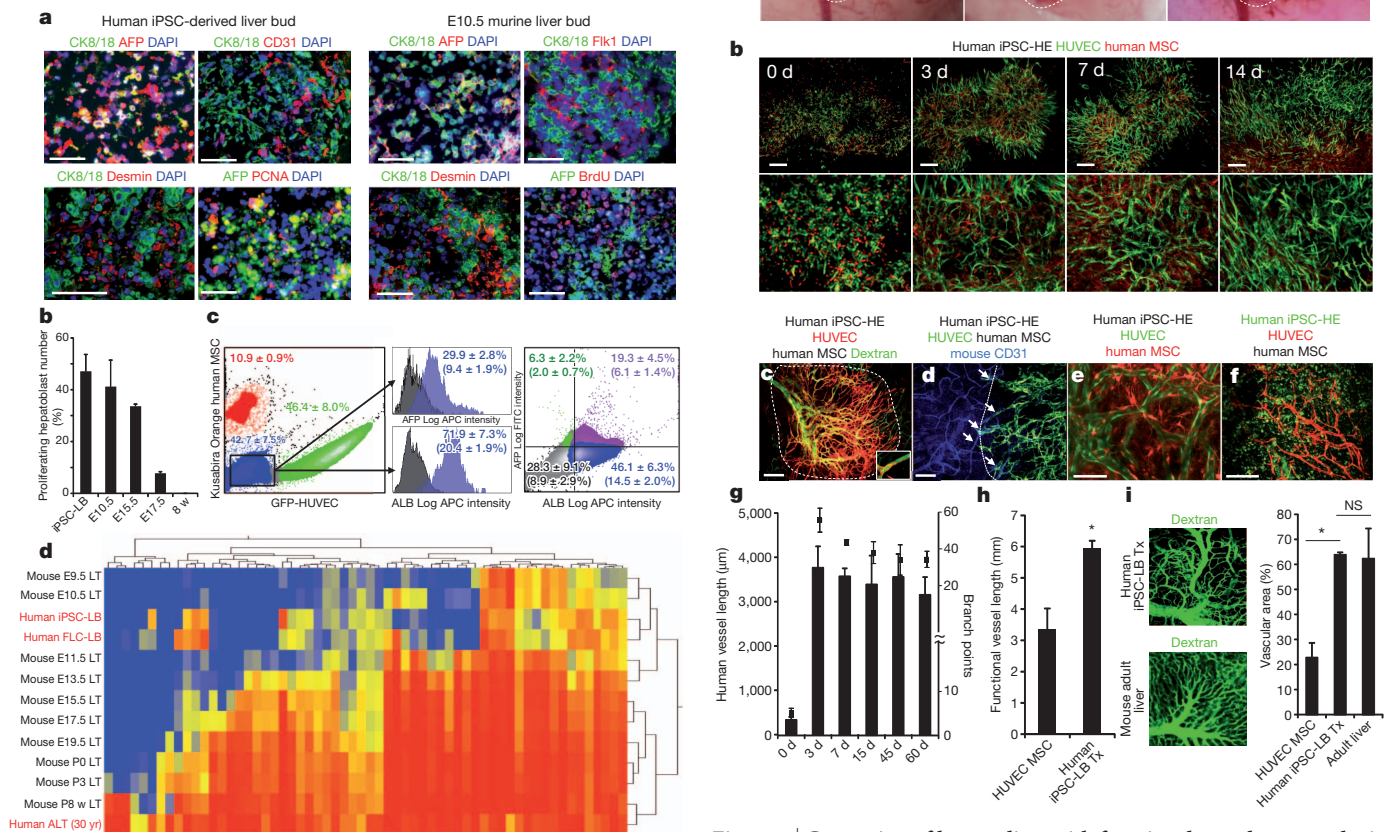


Figure 2 | In vitro characterization of human iPSC-LBs. **a**, Immunostaining of CK8/18, AFP, PECAM1 (CD31), FLK1, desmin, PCNA and BrdU (5-bromodeoxyuridine). Scale bars, 100 μ m. **b**, Proportions of proliferating hepatic cells, as assessed by dividing the number of PCNA-positive or BrdU-positive cells by the number of CK8/18-positive cells (shown as a percentage). Results represent means \pm s.d., $n = 3$. **c**, Representative flow cytometry profile showing the average number of AFP-positive and/or ALB-positive human iPSC-HEs at day 4 of culture in six independent differentiation experiments. Human iPSC-HEs were separated from stromal populations by the use of fluorescence labelled cells. Average percentages of the total cells and s.e.m. are given in brackets. **d**, Comparison of liver developmental gene signatures among human iPSC-LB, human FLC-LB, human adult (30 years old) liver tissue (ALT) and mouse liver tissue (LT) of various developmental stages (from E9.5 to 8 weeks after birth).

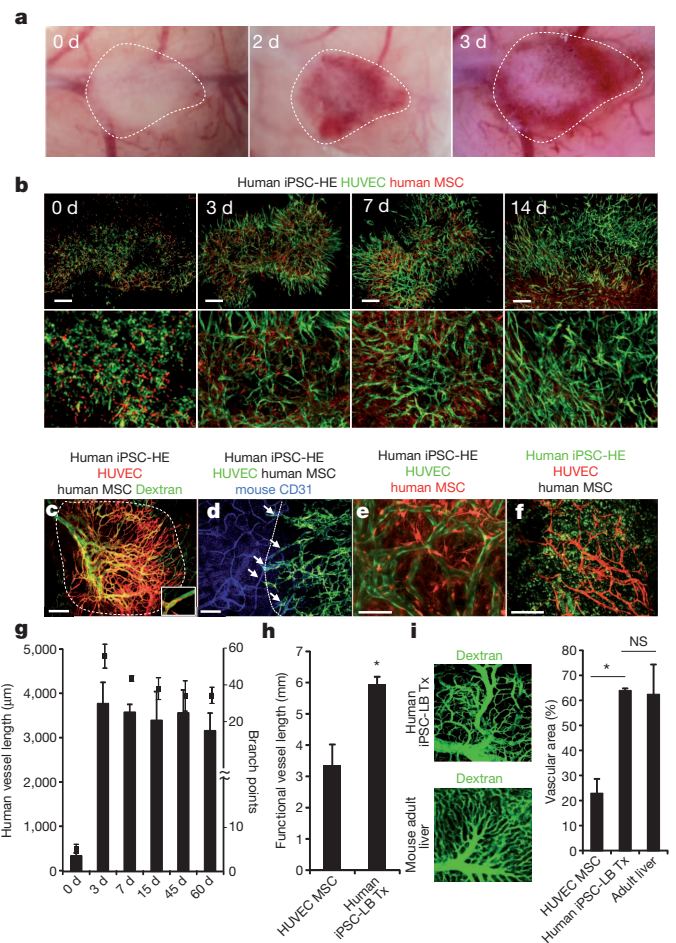


Figure 3 | Generation of human liver with functional vascular networks *in vivo*. **a**, Macroscopic observation of transplanted human iPSC-LBs, showing perfusion of human blood vessels. Dotted area indicates the transplanted human iPSC-LBs. **b**, Intravital tracking of human iPSC-LBs, showing *in vivo* dynamics of vascularization. **c**, Dextran infusion showing the functional human vessel formation at day 3. Scale bar, 500 μ m. **d**, Visualization of the connections (arrows) among HUVECs (green) and host vessels (blue). Scale bar, 250 μ m. **e**, Localization of human MSCs or human iPSC-derived cells at day 15. Scale bars, 100 and 250 μ m. **f**, Localization of human MSCs or human iPSC-derived cells at day 15. Scale bars, 100 and 250 μ m. **g**, Quantification of human vessels over time (mean \pm s.e.m., $n = 3$). Error bars attached to the bars relate to the left axis (vessel length), error bars attached to squares relate to the right axis (branch points). **h**, Comparison of functional vessel length between human iPSC-LB and HUVEC human MSC transplants (Tx) (mean \pm s.e.m., $n = 5$, $*P < 0.01$). **i**, Vascular networks of human iPSC-LB-derived tissue is similar to that of mouse adult livers (mean \pm s.e.m., $n = 5$, $*P < 0.01$).

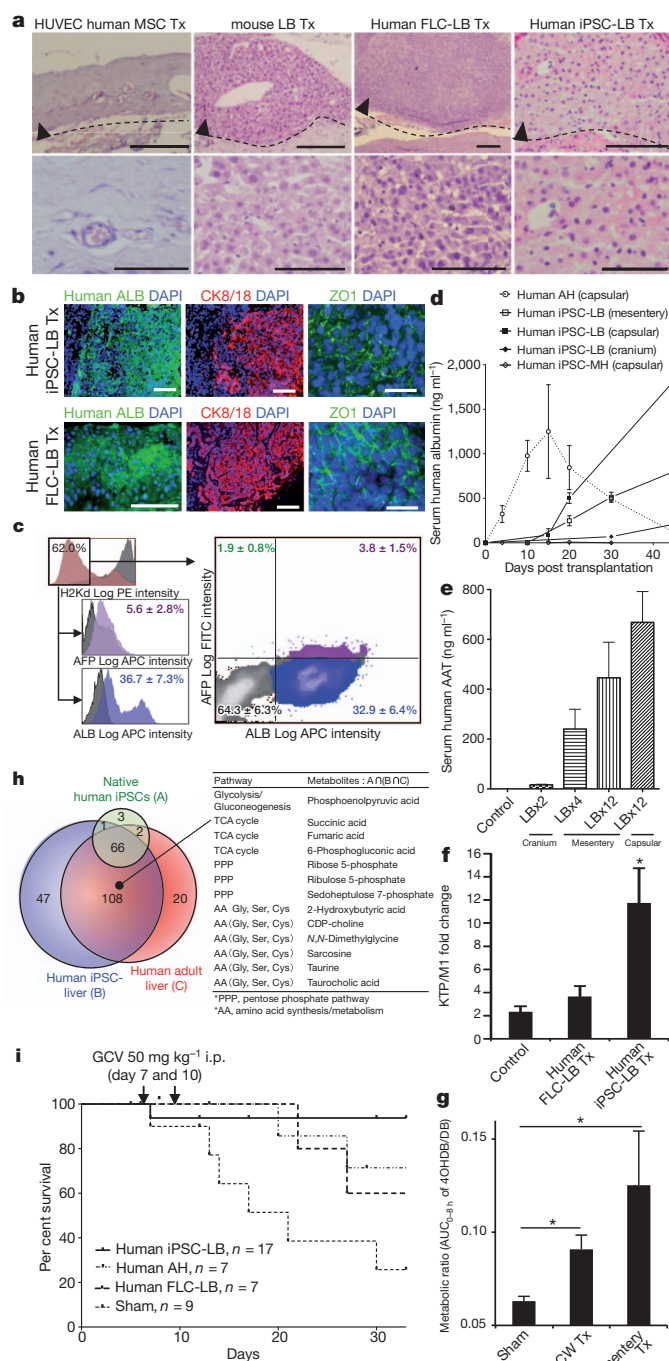


Figure 4 | Functional characterization of human iPSC-LB derived liver.

a, b, Haematoxylin and eosin, and immunofluorescence staining of tissues at day 60. Dotted line indicates the border on the brain. Scale bars, 200 μ m (**a**, top row); 100 μ m (**a**, bottom row and **b**, left and middle images on top and bottom rows) and 25 μ m (**b**, right images, top and bottom rows). **c**, Representative FACS profile for AFP- and/or ALB-positive human cells at day 60 (*n* = 8). Human cells were identified as a murine H2Kd-negative population. Results are shown as contour maps coloured by density of the scatter data. ALB⁻/AFP⁻ population is grey; ALB⁺/AFP⁻ population is green; ALB⁻/AFP⁺ population is blue; and ALB⁺/AFP⁺ population is purple. **d, e**, ELISA showing levels of human serum ALB and AAT over time at various ectopic sites (cranium, mesentery and kidney subcapsule), compared with that of hAH (mean \pm s.e.m., *n* = 4 in **d**, *n* = 3 in **e**). See also Supplementary Tables 2 and 3. **f, g**, Human-specific ketoprofen (**f**) and debrisoquine (**g**) metabolite formations (mean \pm s.e.m., *n* = 3, **P* < 0.05). Metabolic ratios were determined by dividing the AUC_{0-8h} (the area under the curve from time 0 until 8 h; detailed explanation is described in Supplementary Methods) ratio of 4-hydroxydebrisoquine (4OHDB) to debrisoquine (DB). **h**, Venn diagram shows the metabolic profiles of human iPSC-LB transplants measured by CE-TOFMS (capillary electrophoresis time-of-flight mass spectrometry) (Supplementary Fig. 18). Representative metabolites are shown. **i**, Kaplan-Meier survival curves of TK-NOG mice after liver injury (*P* = 0.0120). i.p., intraperitoneal.

mouse CD31 antibody, we showed that human blood vessels within the transplant became patent (unobstructed) by connecting host vessels at the edge of the transplant (Fig. 3c, d), and this was confirmed by whole-mount immunostaining of the explants (Supplementary Figs 9 and 10a, b, and Supplementary Video 4). Functional vessel formation was essential for the successful engraftment because human iPSC-HEs transplanted without endothelial cells failed to vascularize and engraft (Supplementary Fig. 10c–f). Human vessels were long-lasting as they were stabilized by human MSC-derived perivascular cells (Fig. 3e–g)^{13,14}. Interestingly, the human iPSC-LB transplant vascular networks were similar in density and morphology to those of adult livers, whereas the vasculature in only HUVECs and human MSC transplants were much less dense; functional vessels were of similar diameter (approximately 12 μ m) in both settings (in iPSC-LB transplants, and in HUVEC and human MSC transplants) (Fig. 3h, i, and Supplementary Figs 11 and 12). Animal studies suggest that endothelial cells not only form passive conduits to deliver nutrients and oxygen but also establish an instructive vascular niche, which stimulates liver organogenesis and regeneration through elaboration of paracrine trophogens^{5,15}. Our transplantation approach provided a unique intravital monitoring system for evaluating human iPSC-derived organ maturation and differentiation throughout the organogenesis, enabling us to dissect the previously uncharacterized roles of stromal cell types in human organ development.

The liver-bud transplants were examined histologically at multiple time points. Hepatic cells in human iPSC-LBs proliferated rigorously, particularly until 2 months after transplantation (Supplementary Fig. 13). Similar to human FLC-LBs, human iPSC-LB transplants consisted of hepatic cord-like structures characteristic of adult liver after 60 days (Fig. 4a), composed of cells expressing the tight junction protein zona occludens 1 (ZO1), ALB and CK8/18 (Fig. 4b), asialoglycoprotein receptor 1 (ASGR1) and collagen IV (Supplementary Fig. 14), which were normally found along the entire length of the sinusoid¹⁶. AFP-negative transplant-derived hepatic cells had the ultrastructural features of mature hepatocytes (Supplementary Fig. 15a–e). Quantitative fluorescence-activated cell sorting (FACS) profiling of human iPSC-LB transplants showed that 32.9 \pm 6.4% (*n* = 8) of the total human cells were positive for just ALB, whereas 3.8 \pm 1.5% were positive for both ALB and AFP, as seen in human FLC-LB transplants, suggesting the maturation of human iPSC-HEs (Fig. 4c and Supplementary Fig. 16).

To evaluate the functional maturation of human iPSC-LB ectopic transplants compared with that of human adult hepatocytes, human serum albumin was tracked by ELISA. Four transplanted human iPSC-LBs (consisting 1.8 \times 10⁵ immature hepatic cells) began producing albumin at approximately day 10 and produced up to 1,983 ng ml⁻¹ by day 45, and hepatic maturation was observed at iPSC-LBs transplanted at various ectopic sites, whereas 4 \times 10⁶ human adult hepatocytes could not keep producing high levels of albumin and showed a peak at

tissues derived from fetuses and adults 22 to 40 weeks into gestation than did human iPSC-derived mature hepatocyte-like cells (human iPSC-MHs) induced by published protocols⁴. This reflects the efficient hepatic commitment in liver buds as they mimicked ontogenetic interactions consistent with higher albumin production capacity after extended culture (Supplementary Figs 5 and 6).

Haemodynamic stimulation is essential for liver-bud maturation¹². To test whether human iPSC-LBs were capable of generating completely functional liver, we used a cranial window model because of the optical access¹³. We first tested the feasibility of this model by using mouse liver-bud-derived cells (Supplementary Discussion and Supplementary Fig. 7). To track the *in vivo* fate of transplanted cells, we performed repeated live imaging of liver-bud transplant-derived tissue over time. Notably, *in vitro*-derived human iPSC-LBs connected quickly with host vasculature within 48 h of transplantation (Fig. 3a, b, Supplementary Fig. 8 and Supplementary Videos 2 and 3). By infusing fluorescein-conjugated dextran and

day 15 (Fig. 4d). Notably, transplanted conventional human iPSC-MHs (1×10^7 cells per mouse) produced much less albumin, suggesting the importance of three-dimensional and vascularized tissue formation for successful engraftment and maturation at ectopic sites. The amount of human serum alpha-1-antitrypsin (AAT) was in proportion to the number of liver-bud transplants (Fig. 4e). Gene-expression profiling of human iPSC-LBs explanted 60 days ($n = 8$) after transplantation demonstrated the significant hepatic maturation compared with conventional human iPSC-MHs ($n = 8$)⁴ (Supplementary Fig. 17).

To analyse the drug metabolism activity that is a major function of the liver, the mice were challenged with ketoprofen¹⁷ or debrisoquine¹⁸, which are known to be metabolized differently by mice and humans. After the drug exposure, formation of human-specific metabolites was demonstrated in urine and serum samples collected from mice transplanted with human iPSC-LBs (Fig. 4f, g and Supplementary Discussion). Furthermore, to characterize the profiles of small-molecule metabolites such as the products of sugar, amino acid and nucleotide metabolism, metabolome analysis of human iPSC-LB transplants was performed. This showed 222 metabolites, including liver-specific metabolites such as taurocholic acid (Fig. 4h). This profile, with a high number of metabolites, was similar to that of human adult liver but not to that of original human iPSCs (Supplementary Fig. 18). These results demonstrate the potential for predicting drug metabolite profiles of humans, with drug metabolite profiles of mice transplanted with human iPSC-LBs mimicking *in vivo* human physiology. This is particularly striking, as successful detection of human drug responsiveness relies on the use of high-quality hepatocytes and severely damaged host liver using conventional chimaeric mouse models^{19,20}.

With an aim to further our understanding of clinical transplantation and its future use, we evaluated the possibility of a minimally invasive mesenteric transplantation model, which would be a more realistic target site. Macroscopic observation confirmed the successful engraftment of transplanted LBs on mesentery (Supplementary Fig. 19a, b). Transplantation of 12 human iPSC-LBs improved the survival of TK-NOG mice²¹ after the gancyclovir-induced liver failure compared with the sham-operated, human FLC-LB- and human adult-hepatocyte-transplanted mice (Fig. 4i and Supplementary Discussion). Thus, we successfully generated vascularized and functional human liver by transplantation of human iPSCs-LBs.

Manipulation of iPSCs holds great promise for regenerative medicine. However, clinical trials of cell transplantation, currently an important target of the stem-cell-based approach, have presented unsatisfactory results²². Our study demonstrates a proof-of-concept that organ-bud transplantation offers an alternative approach to the generation of a three-dimensional, vascularized organ. These results highlight the enormous therapeutic potential using *in vitro*-grown organ-bud transplantation for treating organ failure.

METHODS SUMMARY

Hepatic differentiation of human iPSCs was achieved based on a protocol reported previously⁴. HUVECs and human MSCs (Lonza) were maintained in endothelial growth medium (Lonza) or MSC growth medium (Lonza) at 37 °C in a humidified 5% CO₂ incubator. To generate human liver buds *in vitro*, 1×10^6 human iPSC-derived hepatic cells, 0.8×10^6 to 1×10^6 HUVECs and 2×10^5 human MSCs were resuspended in the mixture of EGM and hepatocyte culture medium (Cambrex) containing with dexamethasone (0.1 μM, Sigma-Aldrich), oncostatin M (10 ng ml⁻¹, R&D System), hepatocyte growth factor (HGF; 20 ng ml⁻¹, PromoKine) and SingleQuots (Lonza) and plated on pre-solidified $\times 2$ Matrigel diluted with EGM (BD Biosciences) in a 24-well plate. After approximately 4 to 6 days of culture, self-organized human iPSC-LBs were detached, collected and transplanted into a pre-formed cranial window¹³ of an immunodeficient mouse.

Full Methods and any associated references are available in the online version of the paper.

Received 18 April 2012; accepted 7 May 2013.

Published online 3 July 2013.

1. Klein, A. S. *et al.* Organ donation and utilization in the United States, 1999–2008. *Am. J. Transplant.* **10**, 973–986 (2010).

2. Kriks, S. *et al.* Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. *Nature* **480**, 547–551 (2011).
3. Kroon, E. *et al.* Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells *in vivo*. *Nature Biotechnol.* **26**, 443–452 (2008).
4. Si-Tayeb, K. *et al.* Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. *Hepatology* **51**, 297–305 (2010).
5. Matsumoto, K., Yoshitomi, H., Rossant, J. & Zaret, K. S. Liver organogenesis promoted by endothelial cells prior to vascular function. *Science* **294**, 559–563 (2001).
6. Espejel, S. *et al.* Induced pluripotent stem cell-derived hepatocytes have the functional and proliferative capabilities needed for liver regeneration in mice. *J. Clin. Invest.* **120**, 3120–3126 (2010).
7. Kobayashi, T. *et al.* Generation of rat pancreas in mouse by interspecific blastocyst injection of pluripotent stem cells. *Cell* **142**, 787–799 (2010).
8. Zhao, R. & Duncan, S. A. Embryonic development of the liver. *Hepatology* **41**, 956–967 (2005).
9. Si-Tayeb, K., Lemaigre, F. P. & Duncan, S. A. Organogenesis and development of the liver. *Dev. Cell* **18**, 175–189 (2010).
10. Gouysse, G. *et al.* Relationship between vascular development and vascular differentiation during liver organogenesis in humans. *J. Hepatol.* **37**, 730–740 (2002).
11. Jung, J., Zheng, M., Goldfarb, M. & Zaret, K. S. Initiation of mammalian liver development from endoderm by fibroblast growth factors. *Science* **284**, 1998–2003 (1999).
12. Korzh, S. *et al.* Requirement of vasculogenesis and blood circulation in late stages of liver growth in zebrafish. *BMC Dev. Biol.* **8**, 84 (2008).
13. Koike, N. *et al.* Tissue engineering: creation of long-lasting blood vessels. *Nature* **428**, 138–139 (2004).
14. Takebe, T. *et al.* Generation of functional human vascular network. *Transplant. Proc.* **44**, 1130–1133 (2012).
15. Ding, B. S. *et al.* Inductive angiocrine signals from sinusoidal endothelium are required for liver regeneration. *Nature* **468**, 310–315 (2010).
16. Martinez-Hernandez, A. The hepatic extracellular matrix. I. Electron immunohistochemical studies in normal rat liver. *Lab. Invest.* **51**, 57–74 (1984).
17. Ishizaki, T. *et al.* Pharmacokinetics of ketoprofen following single oral, intramuscular and rectal doses and after repeated oral administration. *Eur. J. Clin. Pharmacol.* **18**, 407–414 (1980).
18. Yu, A. M., Idle, J. R. & Gonzalez, F. J. Polymorphic cytochrome P450 2D6: humanized mouse model and endogenous substrates. *Drug Metab. Rev.* **36**, 243–277 (2004).
19. Katoh, M. *et al.* *In vivo* drug metabolism model for human cytochrome P450 enzyme using chimeric mice with humanized liver. *J. Pharm. Sci.* **96**, 428–437 (2007).
20. Kamimura, H. *et al.* Assessment of chimeric mice with humanized liver as a tool for predicting circulating human metabolites. *Drug Metab. Pharmacokinet.* **25**, 223–235 (2010).
21. Hasegawa, M. *et al.* The reconstituted 'humanized liver' in TK-NOG mice is mature and functional. *Biochem. Biophys. Res. Commun.* **405**, 405–410 (2011).
22. Dudley, S. C. Jr. Beware of cells bearing gifts: cell replacement therapy and arrhythmic risk. *Circ. Res.* **97**, 99–101 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank F. Kawamata, E. Yoshizawa, Y. Suzuki, S. Nakai, Y. Takahashi, N. Tsuchida and N. Sasaki for kindly providing technical support; J. Nakabayashi, K. Yasumura, R. Fujiwara, T. Amiya, A. Nakano, Y. Mitsuhashi and all of the members of our laboratory for help with several experiments and comments. We are also grateful to D. Fukumura, Y. Goshima, T. Hirose, M. Ichino, U. Yokoyama, T. Ogawa and R. K. Jain for critical evaluation of the manuscript. This work was supported by grants to H. Taniguchi from the Strategic Promotion of Innovative Research and Development (S-innovation, 62890004) of the Japan Science and Technology Agency (JST). This work was also supported by the Grants-in-Aid of the Ministry of Education, Culture, Sports, Science, and Technology of Japan to T. Takebe (no. 24106510, 24689052), N. Koike (no. 22390260) and H. Taniguchi (no. 21249071, 25253079); by the Specified Research Grant of the Takeda Science Foundation and a grant from the Japan IDDM network to H. Taniguchi; and by a grant of the Yokohama Foundation for Advanced Medical Science to T. Takebe.

Author Contributions T.T. conceived the study, performed the experiments, collected and analysed the data and drafted the manuscript. K.S., M.E., H.K., M.K., T.O., R.-R.Z. and S.A. performed the experiments with the technical guidance and expertise of K.S., Y.-W.Z., Y.U. and T.T. K.S., Y.-W.Z., N.K., Y.A. and H.T. provided critical advice on the research strategy and design.

Author Information Microarray data, including that of human iPSC-LBs, human FLC-LBs, human adult liver tissues (ALT) and mouse liver tissue of various developmental stages, have been deposited in the Gene Expression Omnibus under accession number GSE46631. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.T. (ttakebe@yokohama-cu.ac.jp) or H.T. (rtanigu@yokohama-cu.ac.jp).

METHODS

Cell culture and differentiation. TkDA3 human iPSC clone used in this study was kindly provided by K. Eto and H. Nakauchi. Undifferentiated human iPSCs were maintained on the mouse embryonic fibroblast cells as feeder cells in human iPSC medium. For endodermal differentiation, human iPSCs were seeded on a Matrigel-coated dish and the medium was changed to RPMI1640 with 1% B27 without insulin and human activin A (100 ng ml^{-1}) for 5 to 6 days. For hepatic specification, human iPSC-derived endodermal cells were treated further with RPMI1640 with 1% B27 and human basic FGF (10 ng ml^{-1}), human BMP4 (20 ng ml^{-1}) for 3 to 4 days. Recombinant human activin A was kindly provided by Y. Eto at Ajinomoto Co. Human FLCs (CS-ABI-3716; Applied Cell Biology Research Institute) were plated on collagen IV-coated 6-well plates (BD Biosciences) and cultured in our standard medium (1:1 mixture of DMEM and F-12 (Sigma Aldrich) supplemented with 10% FBS (lot 7219F; ICN Biochemical), 50 mmol l^{-1} HEPES (Wako Pure Chemical Industries), 2 mmol l^{-1} L-glutamine (Life Technologies Corporation), 50 mmol l^{-1} 2-mercaptoethanol (Sigma), $1 \times$ penicillin-streptomycin (Life Technologies), 10 mmol l^{-1} nicotinamide (Sigma), $1 \times 10^{-4} \text{ M}$ Dexamethasone (Sigma) and $1 \mu\text{g ml}^{-1}$ insulin (Wako)). Human recombinant HGF (50 ng ml^{-1}) and EGF (20 ng ml^{-1}) (Sigma) were added before cultivation. HUVECs and human MSCs (Lonza) were maintained in endothelial growth medium or MSC growth medium (Lonza) at 37°C in a humidified 5% CO_2 incubator. Cryopreserved human adult hepatocytes (lots 582, 737, 746 and HC2-8) were purchased from XenoTech (Lenexa). Human adult hepatocytes were thawed, and the number of viable cells was counted according to the manufacturer's instructions and used for transplantation experiments. *In vitro* albumin production tests were performed on viable cells just after first 24 h of culture.

Retroviral transduction. For live imaging, cells were infected with retroviruses expressing EGFP or human Kusabira-Orange (KO1) as described¹³. In brief, a retrovirus vector pGCDNsam IRES-EGFP or KOFP was transfected into 293 gag/pol (gp) and 293gpg (gp and VSV-G) packaging cells (kindly provided by M. Onodera) in which viral particle production is induced using a tetracycline-inducible system. Culture supernatants of retrovirus-infected cells were passed through a $0.45\text{-}\mu\text{m}$ filter (Whatman, GE Healthcare) and used immediately for infection. KOFP displays a major absorption wavelength maximum at 548 nm with a slight shoulder at 515 nm and emits a bright orange fluorescence, with a peak at 561 nm.

Transplantation. *In vitro*-generated liver buds were detached, collected and transplanted into a pre-formed cranial window or indicated sites of a non-obese diabetic/severe combined immunodeficient (NOD/SCID) mouse (Sankyo Lab. Co.). The *in vivo* fate of transplanted cells was monitored by intravital imaging using a fluorescence microscope (model BZ-9000; Keyence) or the Leica TCS SP5 confocal microscope (Leica Microsystems). For survival curves, TK-NOG mice (body weights of approximately 20 to 30 g) were used in this study (supplied by the Central Institute for Experimental Animals)²¹. Ganciclovir (50 mg kg^{-1} , intraperitoneal), a drug that is not toxic to human or mouse tissues, was administered to induce tissue-specific ablation of transgenic liver parenchymal cells at day 7 and 10 after a dozen human iPSC-LBs were transplanted on the mesentery. For control experiments, we thawed and isolated the viable human adult hepatocytes, and resuspended the viable cells (3×10^6 to 4×10^6) in cold medium and cold Matrigel (BD) and transplanted them into the kidney subcapsular space. The mice were bred and maintained according to the Yokohama City University institutional guidelines for the use of laboratory animals.

Intravital imaging. Tail vein injections of 1% tetramethylrhodamine-conjugated dextran (2,000,000 MW), fluorescein-isothiocyanate-conjugated dextran (2,000,000 MW) and Texas-Red-conjugated dextran (70,000 MW, neutral) were used to identify vessel lumens (all from Invitrogen). Host endothelial cells were visualized using Alexa647-conjugated mouse-specific CD31 (BD), injected intravitaly. Confocal image stacks were acquired for the implanted vessels and dextran. Image projections were processed using MetaMorph Angiogenesis Module software (Molecular Devices). Total tubule length, the percentage of tubules per field (area of tubules divided by the area of the transplants that were imaged) and tube diameter were then logged automatically into an Excel spreadsheet.

Gene-expression analysis. Quantitative PCR analyses were conducted as described previously²³. Total RNA of human fetal liver (lot no. A601605) and human adult liver (lot no. B308121) were obtained from Biochain Institute (Hayward).

Microarray and data analysis. Total RNA was prepared from human iPSC-derived cells or tissues (human iPSC-LBs, human FLC-LBs), human adult liver tissue and CD45-negative and Ter119-negative murine liver cells at various different developmental stages (E9.5, E10.5, E11.5, E13.5, E15.5, E17.5, E19.5, postnatal day 0 (P0), P3 and postnatal week 8) using an RNeasy Mini Kit (Qiagen). RNA for gene-expression profiling was hybridized on a Whole Human Genome Agilent $4 \times 44\text{K}$ v2 Oligonucleotide Microarray or Whole Mouse Genome Agilent $4 \times 44\text{K}$ v2 Oligonucleotide Microarray (Agilent Technologies) according to the manufacturer's instructions. To perform cross-species comparison of expression profiles, 26,153 expression data at gene level were cross-referenced to other species using the HomoloGene IDs in the MGI curated data set of human-mouse orthology with Phenotype Annotations (<http://www.informatics.jax.org>). To generate the heat map, we used a hierarchical clustering method with Euclidean distance complete linkage on GeneSpring11.5.1. to analyse the 83 originally selected liver signature gene-expression profiles. Of all genes, 83 genes were selected as liver signature genes because they increased continuously during both murine and human liver development.

ELISA. Blood samples were allowed to clot in a centrifuge tube (approximately 5 min) at room temperature (24°C), loosened from the sides of the tube and kept at 4°C (melting ice) for 20 min. Clotted blood was centrifuged for 10 to 15 min at 400g, 4°C and the serum fraction was removed, taking care to exclude erythrocytes or clotted materials. Human ALB and AAT in the mouse serum samples was measured using a Human Albumin ELISA Quantitation Kit (Bethyl Laboratories) and human alpha 1-antitrypsin ELISA Quantitation Kit (GenWay Biotech) according to the manufacturer's instructions.

Whole mount immunostaining. Mice were perfused with 4% paraformaldehyde in phosphate buffer solution (PBS) through cardiac puncture. The cover-glass forming the cranial window was removed, and the transplants (approximately $300\text{-}\mu\text{m}$ thick) were extracted and placed in 4% paraformaldehyde in PBS for 1.5 h on ice. For immunostaining, fixed collagen gels were washed three times in PBS (10 min each), blocked with 3% BSA or 0.1% Triton X-100 for 1 h, incubated with primary antibodies at 4°C overnight, followed by three 10-min washes in PBS or 0.1% Triton X-100. The sample was incubated with secondary antibodies at 4°C overnight, followed by three 10-min washes in PBS or 0.1% Triton X-100. Tissue samples were counterstained with DAPI and mounted on glass slides in mounting media (Vector Laboratories), under a cover slip. The following primary antibodies were used: mouse anti-human ZO1, mouse anti-human CD31 and rat anti-mouse CD31 (BD Biosciences), rabbit anti-mouse collagen IV (Millipore) and desmin (Dako Corporation). Immunostaining was analysed using the Leica TCS SP5 confocal microscope.

Tissue processing and immunostaining. Tissues were fixed overnight at 4°C in 4% paraformaldehyde, processed, and embedded in paraffin. Transverse sections ($4\text{ }\mu\text{m}$) were placed on MAS-coated slides (Matsunami) for immunostaining or standard histological staining with haematoxylin and eosin. Immunostaining was preceded by autoclave antigen retrieval in citrate buffer (pH 6.0). The primary antibodies were anti-human: CD31, smooth muscle actin, AFP, CK8/18 (all from Dako Corporation) and ALB (BD Biosciences). Tissue sections were incubated with secondary antibody Alexa Fluor (Life Technologies) for 1 h at room temperature, followed by DAPI (Sigma) nuclear staining. The images were acquired using the LSM510 laser scanning microscope (Carl Zeiss).

Statistical analysis. Data are expressed as the means \pm s.d. or s.e.m. from at least three independent experiments. Comparisons between three or four groups were analysed using the Wilcoxon statistical analyses or Kruskal-Wallis test by ranks, and post-hoc comparisons were made using the Mann-Whitney *U*-test with Bonferroni correction. Two-tailed *P* values of <0.05 were considered significant.

23. Kobayashi, S. *et al.* Reconstruction of human elastic cartilage by a CD44⁺ CD90⁺ stem cell in the ear perichondrium. *Proc. Natl Acad. Sci. USA* **108**, 14479–14484 (2011).

Attention enhances synaptic efficacy and the signal-to-noise ratio in neural circuits

Farran Briggs^{1,2}, George R. Mangun^{3,4,5} & W. Martin Usrey^{1,5,6}

Attention is a critical component of perception¹. However, the mechanisms by which attention modulates neuronal communication to guide behaviour are poorly understood. To elucidate the synaptic mechanisms of attention, we developed a sensitive assay of attentional modulation of neuronal communication. In alert monkeys performing a visual spatial attention task, we probed thalamocortical communication by electrically stimulating neurons in the lateral geniculate nucleus of the thalamus while simultaneously recording shock-evoked responses from monosynaptically connected neurons in primary visual cortex. We found that attention enhances neuronal communication by increasing the efficacy of presynaptic input in driving postsynaptic responses, by increasing synchronous responses among ensembles of postsynaptic neurons receiving independent input, and by decreasing redundant signals between postsynaptic neurons receiving common input. The results demonstrate that attention finely tunes neuronal communication at the synaptic level by selectively altering synaptic weights, enabling enhanced detection of salient events in the noisy sensory environment.

Selective attention is a powerful brain mechanism that enables enhanced processing of relevant information while preventing interference from distracting events. Many studies in humans and animals have established that visual attention can influence sensory information processing in visual cortex^{2–8} and subcortical visual areas^{9–11}. Attention directed towards stimuli within the receptive field of a neuron in visual cortex generally results in increases in neuronal firing rate^{12,13} and synchrony^{14,15}. More recent work indicates that visual attention can also alter the correlation structure, variability and/or response gain of neuronal activity^{14,16–18}. However, the fundamental mechanisms by which visual attention alters communication in neural circuits, at the synaptic level, remain a mystery. Moreover, it is unclear how attention-mediated alterations in neuronal population activity translate into improvements in perception¹⁹.

To elucidate the synaptic mechanisms of attention, we developed a sensitive electrophysiological assay of neuronal communication involving stimulation of thalamocortical neurons in the lateral geniculate nucleus (LGN) of the thalamus and simultaneous recordings from monosynaptically connected (that is, postsynaptic) neurons in primary visual cortex (V1) of macaque monkeys performing a spatial attention task. First, we tested whether visual attention alters the efficacy of synaptic communication between the LGN and V1, defined here as the probability that presynaptic stimulation evokes a postsynaptic action potential. Second, we examined whether attention alters both signal and noise in correlated activity among ensembles of postsynaptic target neurons.

Two monkeys were trained to maintain central fixation while covertly focusing their attention on one of two drifting gratings in order to report a contrast change in the attended stimulus (Fig. 1). One of the gratings was positioned over the receptive fields of recorded neurons and the other was located at an equivalent eccentricity away from the

receptive fields. Trials in which attention was directed towards (attend-towards condition) and away (attend-away condition) from the receptive fields of recorded neurons were organized into blocks and cued by the colour of the central fixation dot. In a random 5% of the trials the cue instruction was invalid, such that the contrast change occurred at the unattended location. Animals were rewarded for correct detection of the contrast change in validly and invalidly cued trials. Behavioural measures of spatial attention were derived by comparisons of accuracy (percentage of trials completed correctly) and reaction times in validly and invalidly cued trials. For both monkeys, accuracy was significantly greater ($P < 0.03$) and reaction times were significantly faster ($P < 0.05$; Fig. 1b) for validly versus invalidly cued trials, indicating that animals were covertly attending to the specified location.

In each animal, we implanted stimulating electrodes in the magnocellular and parvocellular layers of the LGN (Fig. 2a), so that weak electrical shocks applied to thalamocortical neurons in these layers evoked suprathreshold, short- and fixed-latency monosynaptic spikes in recorded (postsynaptic) thalamocortical-recipient (TCR) neurons, located in layer 4C of V1 (Fig. 2a, b). Importantly, stimulation levels were set so that stimulation evoked a postsynaptic spike in only a fraction of trials (Supplementary Fig. 1a). We recorded visually evoked activity in response to drifting sinusoidal gratings in order to characterize the physiological responses of all recorded TCR neurons. TCR neurons ($n = 61$) were grouped into those receiving input from the magnocellular layers and from the parvocellular layers (we refer to these as magnocellular-recipient ($n = 36$) and parvocellular-recipient ($n = 25$) populations, respectively) based on the stimulus contrast required to evoke a half-maximum response (Fig. 2c). Magnocellular- and parvocellular-recipient neurons differed across several physiological

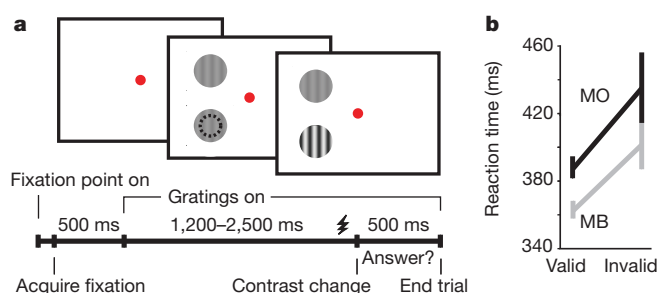
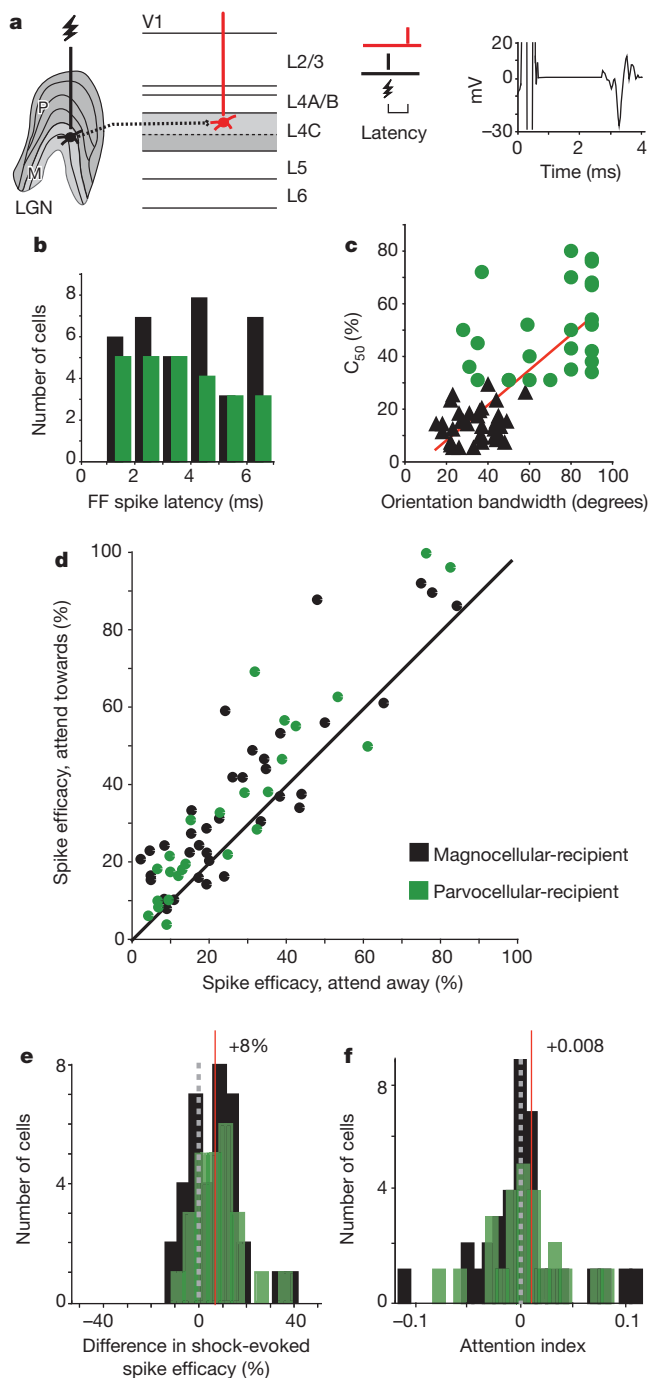


Figure 1 | Attention task and behavioural performance. **a**, The attention task, including representative frames of the visual display for a validly cued trial in which attention was directed by cue colour (the central fixation point; red in this case) towards the receptive field of the neuron. Dashed black circle (middle frame) represents the receptive field of the recorded neuron. The timeline for one trial is shown at bottom; LGN shock timing is indicated schematically just before the contrast change (shocks occurred on 70% of trials). **b**, Reaction time data for validly versus invalidly cued trials for the two monkeys (Monkey B (MB), $P < 0.01$; Monkey O (MO), $P < 0.05$). Error bars represent s.e.m.

¹Center for Neuroscience, University of California, Davis, 1544 Newton Court, Davis, California 95618, USA. ²Department of Physiology and Neurobiology, Geisel School of Medicine at Dartmouth, 1 Medical Center Drive, Lebanon, New Hampshire 03756, USA. ³Center for Mind and Brain, University of California, Davis, 267 Cousteau Place, Davis, California 95618, USA. ⁴Department of Psychology, University of California, Davis, 1 Shields Avenue, Davis, California 95616, USA. ⁵Department of Neurology, University of California, Davis, 4860 Y Street, Sacramento, California 95817, USA. ⁶Department of Neurobiology, Physiology and Behavior, University of California, Davis, 1 Shields Avenue, Davis, California 95616, USA.



parameters, including orientation tuning ($P < 0.001$; Fig. 2c), the ratio of the first harmonic (f1; fundamental frequency) to mean response (f0; $P < 0.001$), and visually evoked firing rates ($P < 0.05$). However, shock efficacies (Supplementary Fig. 1a), spontaneous firing rates, and shock-evoked postsynaptic-response latencies (Fig. 2b) did not vary significantly across the magnocellular- and parvocellular-recipient populations, consistent with previous reports²⁰.

After TCR neurons were characterized physiologically, we recorded both visually evoked and shock-evoked neuronal responses while animals performed the attention task. In a subset of attention trials (70%), a single shock was delivered to the LGN between 1,000 and 1,200 ms after the onset of grating presentation, always at the same phase of the visual stimulus cycle, and before the contrast change. Electrical stimulation did not affect performance, as there was not a significant difference in the animals' ability to complete shock and non-shock trials in

Figure 2 | Attentional modulation of thalamocortical synaptic efficacy.

a, Experimental setup. Electrical stimulation of presynaptic LGN neurons (black) leads to a postsynaptic response in the simultaneously recorded TCR neuron (red). Shock-evoked postsynaptic spikes occur at fixed latencies with little temporal jitter, as illustrated by the trial-averaged waveform of a representative TCR neuron (right). **b**, Distribution of postsynaptic-response latencies for magnocellular- and parvocellular-recipient neurons in black and green, respectively (mean latency; magnocellular-recipient neurons = 3.4 ± 0.3 ms, parvocellular-recipient neurons = 3.1 ± 0.4 ms). FF, feedforward. **c**, Relationship between C_{50} (contrast for half-maximum response) and orientation-tuning bandwidth (peak half-width, at half height) for magnocellular- and parvocellular-recipient neurons ($R^2 = 0.52$). Magnocellular-recipient neurons; $C_{50} = 13.8 \pm 1.1\%$, orientation half-width = $34 \pm 1.8^\circ$. Parvocellular-recipient neurons; $C_{50} = 49.4 \pm 3.4\%$, orientation half-width = $69 \pm 4.6^\circ$. Colour conventions as in **b**. **d**, Percentage of shock-evoked postsynaptic spikes in attend-towards versus attend-away conditions. Black line represents unity. Average efficacy (percentage of shocks that evoke a postsynaptic response): for all TCRs, attend towards = $36 \pm 3\%$, attend away = $28 \pm 3\%$; magnocellular-recipient neurons, attend towards = $37 \pm 4\%$, attend away = $29 \pm 4\%$; parvocellular-recipient neurons, attend towards = $35 \pm 5\%$, attend away = $28 \pm 5\%$. **e**, **f**, Distribution of differences in (e) shock-evoked spike efficacy (attend-towards versus attend-away condition) and (f) attention-index values for magnocellular- and parvocellular-recipient neurons (note difference in scales). Attention index values calculated from firing rate 850 to 1,200 ms after onset of grating stimulation. Dashed lines indicate zero and red lines indicate mean values (mean difference in spike efficacy for magnocellular- and parvocellular-recipient TCRs = $8 \pm 2\%$; mean attention index magnocellular-recipient = 0.008 ± 0.007 , parvocellular-recipient = 0.006 ± 0.007 ; colour conventions as in **b**).

the attend-towards or attend-away conditions ($P > 0.5$). When we compared thalamocortical synaptic efficacy (per cent of presynaptic shocks that evoked a postsynaptic spike) as a function of attention, we observed a highly significant increase in synaptic efficacy when covert spatial attention was directed towards the receptive fields of recorded TCR neurons compared to when attention was directed away from TCR receptive fields ($P < 0.001$; Fig. 2d). Attentional modulation of synaptic efficacy was significant for both magnocellular- and parvocellular-recipient neuronal populations ($P < 0.001$ for both). For the attend-away trials, it is interesting to note that the number of spikes occurring within the time window corresponding to the postsynaptic response latency did not differ between shock and non-shock trials (Supplementary Fig. 1b), despite the fact that shock strength was set to be effective in neutral conditions. This finding is consistent with the view that attention may also have a suppressive influence on synaptic efficacy when directed away from a neuron's receptive field.

The robust attentional enhancement of thalamocortical synaptic efficacy, indexed by the positive shift in shock-evoked spike efficacy across attention conditions (Fig. 2e), contrasts with a modest attention-mediated increase in neuronal firing rate (Fig. 2f; statistically significant for magnocellular-recipient ($P < 0.025$), but not parvocellular-recipient ($P > 0.1$) neurons). Interestingly, this modest increase in firing rate across attention conditions was only evident for later time periods between stimulus onset and the earliest opportunity for contrast change (850 to 1,200 ms and 1,000 to 1,200 ms), and was not evident over earlier or broader time periods (0 to 1,200 ms or 600 to 1,200 ms; $P > 0.75$; Supplementary Fig. 1c), consistent with previous studies showing weak or no attentional modulation of firing rate in V1 (refs 7, 21–24). Further analysis also demonstrated that there was no relationship between the magnitude of attention effects on firing rate and the influence of attention on the efficacy of thalamocortical communication (Supplementary Fig. 1d). These results suggest that; first, attentional modulation of synaptic efficacy in thalamocortical circuits is not merely due to simple gain or firing-threshold changes in LGN or TCR neurons; and second, that thalamocortical visual pathways may make use of a different (synaptic-level) mechanism to propagate attentional signals compared to higher visual areas, where attentional modulation can be indexed by changes in neuronal firing rates.

We explored further the temporal precision of attentional modulation of synaptic efficacy in thalamocortical circuits by examining spiking

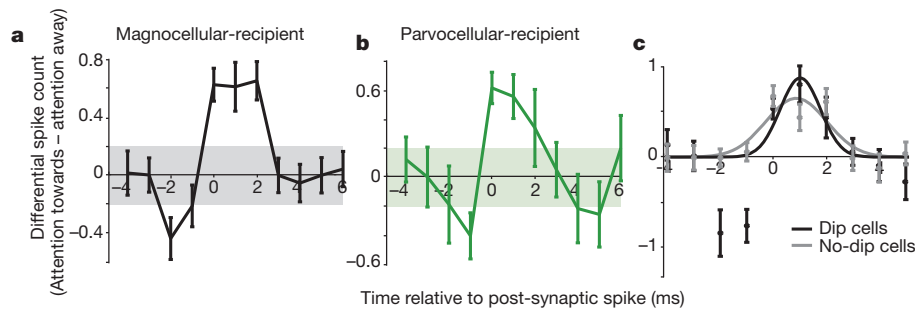


Figure 3 | Temporal precision of attentional enhancement of synaptic efficacy. **a, b,** Average differential spiking activity (attend towards—attend away) surrounding the time of the shock-evoked postsynaptic spike (occurring at time = 0) for magnocellular-recipient TCR neurons (**a**), and parvocellular-recipient TCR neurons (**b**). Error bars represent s.e.m.; shaded

activity over a 10-ms window surrounding the time of the shock-evoked postsynaptic spike (−4 ms to +6 ms). Figure 3a, b illustrates the population average time-course of differential spiking activity (attend towards—attend away) surrounding the shock-evoked postsynaptic spike (defined as time = 0) for magnocellular- and parvocellular-recipient neurons, respectively (see Supplementary Fig. 2a for separate time-course plots corresponding to each attention condition). In both cases, spiking activity rises above two standard deviations of the mean activity for 2 to 3 ms, indicating that attention causes an increase in the number of synaptically evoked spikes during a limited time window. Notably, about 40% of TCR neurons (42% of magnocellular-recipient neurons, and 40% of parvocellular-recipient neurons) displayed a prominent dip in spiking activity just before the evoked spike (Fig. 3a, b and Supplementary Fig. 2a, b), suggesting that fast feedforward inhibition, probably via local interneurons, may suppress spiking activity just before the occurrence of the postsynaptically evoked spike. Consistent with this, we found that TCR subpopulations with fast feedforward inhibition showed less jitter in the timing of their shock-evoked postsynaptic spikes (response profile = 1.75 ms for ‘dip’ neurons versus 2.75 ms for ‘no-dip’ neurons; Fig. 3c). These results suggest that feedforward inhibition sharpens postsynaptic spike-timing precision by approximately 1 ms. The presence of dips illustrates that attentional modulation in our sample of V1 neurons was often dynamic, including phases of reduced spiking as well as phases of increased spiking. These dynamics could help to explain why attentional modulation did not produce a larger increase in overall firing rate.

After it was established that attention alters the efficacy of synaptic communication in neural circuits with fine temporal precision, we next set out to determine whether attention differentially affects the processing of signal and noise in thalamocortical circuits. Previous studies have provided evidence to suggest that attention may alter neuronal activity by increasing signal or reducing noise in firing-rate fluctuations^{20,22}. We assessed whether attention could both boost the signal and reduce noise in the same circuit by examining the effects of attention on the occurrence of synchronized spikes between pairs of simultaneously recorded TCR neurons. In some circumstances, synchronized spiking may boost signal detection by driving downstream targets more effectively^{25,26}, whereas in other cases, synchronized spiking may confound decoding by downstream target neurons through the introduction of noise²⁷. Using a multi-electrode recording array, we recorded from 71 pairs of simultaneously recorded TCR neurons that fired synchronized spikes in response to electrical stimulation in the LGN. Our criteria for identifying synchronous postsynaptic spikes were strict: in any given trial, synchronous spikes needed to occur at the specified shock-evoked postsynaptic spike latencies for each TCR neuron in the pair. For each pair, we calculated the percentage of trials in which synchronous spikes were evoked in response to electrical stimulation in the attend-towards versus attend-away conditions. Results from this analysis were clear; attention significantly increased the percentage

regions represent 2 s.d. above and below mean activity. **c**, Average differential spiking activity of TCR neurons separated into groups on the basis of whether they displayed early inhibition (dip versus no-dip cells). Error bars represent s.e.m. Black and grey lines illustrate Gaussian fits to dip and no-dip cell data. Width at half height for dip cells = 1.75 ms; for no-dip cells = 2.75 ms.

of synchronous spikes across our sample of TCR pairs (Fig. 4a, b; $P < 0.001$). This effect was present for magnocellular–magnocellular, parvocellular–parvocellular, and magnocellular–parvocellular pairs in our sample ($P < 0.04$; Fig. 4a, b, black, green and grey symbols, respectively), suggesting that attentional modulation of spike synchrony is consistent across thalamocortical circuits.

Synchronized spiking between pairs of TCR neurons could arise from two sources: first, the simultaneous arrival of spikes travelling in independent channels of communication (for example, different LGN axons); and second, the simultaneous arrival of spikes travelling in a common channel (for example, one LGN axon) with divergence. Although both mechanisms will propagate signal (stimulus-evoked) and noise spikes to their postsynaptic targets, randomly generated noise spikes are less likely to occur simultaneously between independent LGN axons than between the branches of common input LGN axons. Consequently, attention could increase the ratio of signal-to-noise in thalamocortical circuits by increasing the strength of spikes arriving from independent channels, and also by reducing the strength of spikes arriving from common-input channels. To determine whether pairs of TCR neurons in our sample received common LGN input, we calculated shuffle-corrected cross-correlograms from TCR neuronal responses to drifting sinusoidal gratings. For 25 TCR pairs, cross-correlograms contained a single, narrow (< 3 ms, full width at half height) peak centred at time zero, indicating that the two neurons frequently fired synchronous spikes in response to common feedforward input (Fig. 4c). TCR pairs in our sample with zero-centred cross-correlogram peaks had overlapping receptive fields, consistent with previous studies of correlation patterns shown by visual neurons with common feedforward input²⁶. Most recorded TCR pairs consisted of two magnocellular-recipient neurons or two parvocellular-recipient neurons, consistent with the anatomical segregation of magnocellular and parvocellular inputs to V1. However, we also encountered pairs of mixed magnocellular- and parvocellular-recipient neurons.

For TCR pairs receiving common input, we calculated the percentage of synchronous spikes (that is, the percentage of spikes contained in the cross-correlogram peak) when attention was directed towards and away from the cells’ receptive fields. Results of this analysis showed that attention decreased the percentage of synchronous spikes from common input by approximately 10% (see Supplementary Methods). Consistent with this, the distribution of differential cross-correlogram peak height (attend towards—attend away) was shifted significantly to the left of zero ($P < 0.01$), indicating that attention decreased synchronous responses to common input (Fig. 4d). These results suggest that attention may decrease noise in thalamocortical communication by reducing the amount of synchronous spikes arising from common feedforward input.

As a final analysis, we explored whether attention differentially modulated synchronous spiking in pairs of TCR neurons receiving input from independent sources (that is, TCR pairs with flat cross-correlograms)

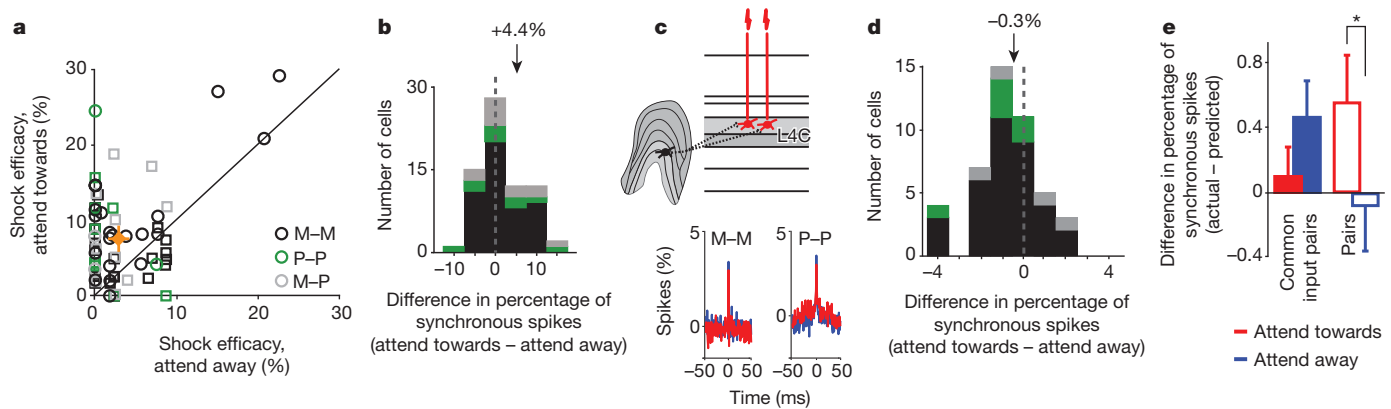


Figure 4 | Attentional modulation of synchronized spiking. **a**, The efficacy of shocks in evoking synchronous spikes; the percentage of synchronously evoked postsynaptic spikes across attention conditions for 71 pairs of simultaneously recorded magnocellular–magnocellular (M–M; black, $n = 48$), parvocellular–parvocellular (P–P; green, $n = 11$), magnocellular–parvocellular (M–P; grey, $n = 12$) TCR pairs (circles represent TCR pairs; squares represent putative TCR pairs). Black line represents unity. Average efficacy of synchronous spiking across all pairs in the attend-towards condition = $7.6 \pm 0.8\%$, and in the attend-away condition = $3.1 \pm 0.6\%$ (orange diamond, cross-hairs represent s.e.m.). **b**, Distribution of attention-mediated differences in the percentage of shock-evoked synchronous spikes. Colour conventions as in **a**. Dashed line represents zero and arrow illustrates the population mean ($+4.4 \pm 0.7\%$ s.e.m.). **c**, Diagram illustrating a pair of TCR neurons receiving common presynaptic input and two examples of shuffle-corrected cross-correlograms for magnocellular–magnocellular-recipient and parvocellular–parvocellular-recipient pairs, illustrating the occurrence of

synchronous spikes (narrow peaks centred at time zero) and the influence of attention on the percentage of synchronous spikes in attend-towards (red) and attend-away (blue) conditions. **d**, Distribution of attention-mediated differences in correlated spikes among pairs receiving common input (21 magnocellular–magnocellular, 2 parvocellular–parvocellular, and 2 magnocellular–parvocellular pairs). Conventions as in **b**. Mean difference in spikes in peak = $-0.3 \pm 0.2\%$. **e**, Difference between actual and predicted synchronous spikes for TCR pairs receiving common feedforward input (solid bars; $n = 25$) and TCR pairs receiving independent input (open bars; $n = 46$) across attention conditions. Error bars represent s.e.m.; asterisk indicates significant difference across attention conditions for TCR pairs receiving independent input ($P < 0.025$). Average actual – predicted values: common input TCR pairs, attend towards = $0.1 \pm 0.2\%$, attend away = $0.5 \pm 0.2\%$; independent input TCR pairs, attend towards = $0.6 \pm 0.3\%$, attend away = $-0.1 \pm 0.3\%$.

compared to TCR pairs that received input from common sources (that is, TCR pairs with narrow, zero-centred cross-correlogram peaks). More specifically, we determined whether measured percentages of synchronous spikes differed from predicted percentages based on the product of the evoked postsynaptic spike probabilities measured for each TCR neuron in the pair (Fig. 4e). For TCR neurons receiving common input, the percentage of measured synchronous spikes was not significantly different from the prediction ($P > 0.25$), and there was no effect of attention on this relationship. However, for TCR pairs receiving independent input, there was a significant difference between actual and predicted synchronous spikes, and attention had a significant effect on the relationship between measured and predicted values ($P < 0.025$). These results suggest that attention differentially regulates synchronized inputs emerging from independent and common sources. Accordingly, attention boosts signal transmission by enhancing responses to synchronous inputs from independent sources and reduces noise transmission by reducing responses to synchronous inputs from common input sources. Based on our data, attention may boost signal-to-noise ratios on average by approximately 20% (see Methods). This finding has significant implications for understanding the mechanisms by which neural networks optimally encode sensory information in the face of potentially noisy correlations resulting from anatomical wiring constraints²⁷.

Taken together, the results of this study provide multiple insights into the mechanisms by which attention alters neuronal communication. First, attention modulates signal transmission by enhancing synaptic efficacy. This finding represents the first evidence that attention acts at the synaptic level. Second, attention modulates afferent signal transmission with fine temporal precision. Third, attention serves to increase the ratio of signal to noise in neural circuits by simultaneously enhancing the transmission of signal and reducing the transmission of noise. These results suggest strongly that attention modulates synaptic inputs in a highly selective manner, such that inputs that carry salient sensory information (through independent channels) are enhanced and inputs carrying potentially redundant information (through common channels) are suppressed. Each of these results has significant

implications for our understanding of attentional modulation of sensory information processing.

Attentional modulation of synaptic efficacy in thalamocortical circuits was robust and displayed temporal precision. Attention-related improvements in spike-timing precision in V1 resulted in part from fast disinhibitory feedforward inhibition. Interestingly, the temporal precision of attentional modulation of V1 activity did not correlate strongly with more global changes in firing rate (represented by an attention index calculated from peri-stimulus spiking activity). This lack of correspondence suggests that attention alters brain activity through multiple mechanisms, including more global alterations in neuronal firing rate, as well as finer-scale dynamic alterations in synaptic communication operating at the level of individual neural circuits. Moreover, our results support the idea that attentional modulations involving fine-scale dynamics may not manifest in more global alterations in neuronal firing rate. At the local circuit level, this effect may serve to enhance spatial and temporal precision, but at the more global level, these effects may average out. In V1 (and other sensory cortices), attention may make use of fine-scale dynamics to accommodate depressing synapses, a known property of thalamocortical afferents²⁸. It would be interesting to know whether or not attention affects synaptic weights in higher visual cortical areas and, if so, whether the effects of attention on synaptic weights underlie the influence of attention on neuronal firing-rate dynamics.

Our data support the idea that attention enhances sensory information processing directly by increasing the ratio of signal to noise in neural-circuit communication. Simultaneous signal enhancement and noise reduction in the same neural circuit suggests that attention modulates correlated synaptic activity in a highly selective manner. Select synaptic connections that originate from independent inputs and carrying feature-specific information about a sensory stimulus are more strongly weighted with attention, leading to better processing of salient stimulus features by downstream neurons. Synaptic connections that originate from common inputs are weighted less with attention, so that false positives are less likely to be communicated to downstream decoding neurons. Such asymmetric synaptic weighting with

attention hints at a presynaptic locus for modulation, because a post-synaptic locus, such as altering the membrane potential threshold of cortical recipient neurons, would be difficult to reconcile with asymmetric synaptic weights. Furthermore, the finding that attention does not increase the overall firing rate of cortical neurons that receive direct LGN input indicates that the measured changes in thalamocortical communication are unlikely to be due to a generalized depolarization among target neurons with attention. To determine the structural basis for presynaptic modulation is beyond the scope of the current study, but one possible candidate is differential modulation by acetylcholine. Acetylcholine has been implicated in attention effects in V1 (ref. 29), and a particular class of cholinergic receptors are localized to the presynaptic terminals of LGN axons that innervate cortical layer 4C neurons³⁰. These cholinergic synapses could therefore provide a route for attention to alter synaptic weights selectively.

Feedforward subcortical–cortical and cortico–cortical connections often must convey information with speed and precision, but anatomical wiring constraints on these connections can introduce unreliable information. Here we demonstrate that attention alters synaptic communication in a dynamic and highly selective manner that could be uniquely adapted for signal transmission in sensory cortex. Specifically, attention selectively enhances inputs carrying salient sensory information while simultaneously suppressing inputs carrying potentially redundant information. These findings suggest that attention could represent a critical mechanism by which anatomical wiring limitations are overcome in order to optimize communication across neural circuits, thereby permitting the most behaviourally relevant information to influence perception and performance.

METHODS SUMMARY

Two adult female macaque monkeys (*Macacca mulatta*) were used for this study. All of the procedures carried out as a part of this study conformed to the guidelines set by the National Institutes of Health and were approved by the Institutional Animal Care and Use Committee at the University of California, Davis. Under full surgical anaesthesia, animals were implanted with head-posts and cylinders encircling two craniotomies over the LGN and V1. Animals learned a contrast-change detection task requiring covert shifts in visual spatial attention. Stimulating electrodes were semi-chronically implanted within the LGN and recording electrodes were placed in retinotopically aligned regions of V1. V1 neurons monosynaptically connected to LGN inputs were identified by short, fixed-latency action potentials after electrical stimulation of LGN neurons. Responses of (postsynaptic) thalamocortical-recipient (TCR) neurons in V1 to visual stimulation during trials in which animals' directed attention towards or away from drifting sinusoidal gratings placed within recorded neurons' receptive fields were compared. Electrical shocks were delivered to the LGN on a proportion of attention trials and the probability of a shock evoking a spike in TCR neurons was compared across attention conditions. Single-unit recordings of TCR neurons were made with single electrodes and a five-channel multi-electrode array. Spikes were sorted offline using principle components analysis. Subsequent data analyses included calculations of tuning responses in order to classify TCR neurons as receiving magnocellular or parvocellular LGN input, calculations of attention-index values, determination of shock-evoked spiking probability, and calculation of shuffle-corrected cross-correlograms for all simultaneously recorded pairs of TCR neurons.

Full Methods and any associated references are available in the online version of the paper.

Received 15 February; accepted 9 May 2013.

Published online 26 June 2013.

1. Posner, M. I., Synder, C. R. R. & Davidson, B. J. Attention and the detection of signals. *J. Exp. Psychol.* **109**, 160–174 (1980).
2. Van Voorhis, S. & Hillyard, S. A. Visual evoked potentials and selective attention to points in space. *Percept. Psychophys.* **22**, 54–62 (1977).
3. Moran, J. & Desimone, R. Selective attention gates visual processing in the extrastriate cortex. *Science* **229**, 782–784 (1985).

4. Heinze, H. J. *et al.* Combined spatial and temporal imaging of brain activity during visual selective attention in humans. *Nature* **372**, 543–546 (1994).
5. Ito, M. & Gilbert, C. D. Attention modulates contextual influences in the primary visual cortex of alert monkeys. *Neuron* **22**, 593–604 (1999).
6. Kelly, S. P., Gomez-Ramirez, M. & Foxe, J. J. Spatial attention modulates initial afferent activity in human primary visual cortex. *Cereb. Cortex* **18**, 2629–2636 (2008).
7. McAdams, C. J. & Reid, R. C. Attention modulates the responses of simple cells in monkey primary visual cortex. *J. Neurosci.* **25**, 11023–11033 (2005).
8. Thiele, A., Pooresmaeili, A., Delicato, L. S., Herrero, J. L. & Roelfsema, P. R. Additive effects of attention and stimulus contrast in primary visual cortex. *Cereb. Cortex* **19**, 2970–2981 (2009).
9. Vanduffel, W., Tootell, R. B. & Orban, G. A. Attention-dependent suppression of metabolic activity in the early stages of the macaque visual system. *Cereb. Cortex* **10**, 109–126 (2000).
10. O'Connor, D. H., Fukui, M. M., Pinsk, M. A. & Kastner, S. Attention modulates responses in the human lateral geniculate nucleus. *Nature Neurosci.* **5**, 1203–1209 (2002).
11. McAlonan, K., Cavanaugh, J. & Wurtz, R. H. Guarding the gateway to cortex with attention in visual thalamus. *Nature* **456**, 391–394 (2008).
12. Spitzer, H., Desimone, R. & Moran, J. Increased attention enhances both behavioral and neuronal performance. *Science* **240**, 338–340 (1988).
13. McAdams, C. J. & Maunsell, J. H. Attention to both space and feature modulates neuronal responses in macaque area V4. *J. Neurophysiol.* **83**, 1751–1755 (2000).
14. Fries, P., Reynolds, J. H., Rorie, A. E. & Desimone, R. Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* **291**, 1560–1563 (2001).
15. Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I. & Schroeder, C. E. Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* **320**, 110–113 (2008).
16. Cohen, M. R. & Maunsell, J. H. Attention improves performance primarily by reducing interneuronal correlations. *Nature Neurosci.* **12**, 1594–1600 (2009).
17. Mitchell, J. F., Sundberg, K. A. & Reynolds, J. H. Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* **63**, 879–888 (2009).
18. Pestilli, F., Carrasco, M., Heeger, D. J. & Gardner, J. L. Attentional enhancement via selection and pooling of early sensory responses in human visual cortex. *Neuron* **72**, 832–846 (2011).
19. Zénon, A. & Krauzlis, R. J. Attention deficits without cortical neuronal deficits. *Nature* **489**, 434–437 (2012).
20. Bullier, J. & Henry, G. H. Ordinal position and afferent input of neurons in monkey striate cortex. *J. Comp. Neurol.* **193**, 913–935 (1980).
21. Chen, Y. *et al.* Task difficulty modulates the activity of specific neuronal populations in primary visual cortex. *Nature Neurosci.* **11**, 974–982 (2008).
22. Luck, S. J., Chelazzi, L., Hillyard, S. A. & Desimone, R. Neural mechanisms of spatial selective attention in area V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* **77**, 24–42 (1997).
23. Motter, B. C. Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J. Neurophysiol.* **70**, 909–919 (1993).
24. Yoshor, D., Ghose, G. M., Bosking, W. H., Sun, P. & Maunsell, J. H. R. Spatial attention does not strongly modulate neuronal responses in early human visual cortex. *J. Neurosci.* **27**, 13205–13209 (2007).
25. Usrey, W. M. The role of spike timing for thalamocortical processing. *Curr. Opin. Neurobiol.* **12**, 411–417 (2002).
26. Usrey, W. M., Alonso, J.-M. & Reid, R. C. Synaptic interactions between thalamic inputs to simple cells in cat visual cortex. *J. Neurosci.* **20**, 5461–5467 (2000).
27. Zohary, E., Shadlen, M. N. & Newsome, W. T. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* **370**, 140–143 (1994).
28. Stratford, K. J., Tarczy-Hornoch, K., Martin, K. A. C., Bannister, N. J. & Jack, J. J. B. Excitatory synaptic inputs to spiny stellate cells in cat visual cortex. *Nature* **382**, 258–261 (1996).
29. Herrero, J. L. *et al.* Acetylcholine contributes through muscarinic receptors to attentional modulation in V1. *Nature* **454**, 1110–1114 (2008).
30. Disney, A. A., Aoki, C. & Hawken, M. J. Gain modulation by nicotine in macaque V1. *Neuron* **56**, 701–713 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank K. E. Neverkovic, D. J. Sperka and R. Oates-O'Brien for technical and veterinary assistance. This work was supported by National Institutes of Health grants EY18683 (F.B.), EY013588 (W.M.U.), MH055714 (G.R.M.) and NSF grants BCS-0727115 and 1228535 (G.R.M. and W.M.U.).

Author Contributions F.B., G.R.M. and W.M.U. designed the experiments. F.B. conducted the experiments and performed the data analyses in collaboration with G.R.M. and W.M.U. F.B., G.R.M. and W.M.U. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.M.U. (wrmusrey@ucdavis.edu).

METHODS

Two adult female macaque monkeys (*Macacca mulatta*) were used for this study. All of the procedures performed as part of this study conformed to the guidelines set by the NIH and were approved by the Institutional Animal Care and Use Committee at the University of California, Davis.

Surgical preparation. All surgical procedures have been described in detail previously^{31,32}. Under full surgical anaesthesia, two craniotomies were made to enable recording access to the LGN and to the parafoveal opercular region of V1. Two recording cylinders were placed encircling the craniotomies and incased in an implant of bone cement. Head-restraint posts were also attached to cranial implants. After recovery from surgery, cylinders were flushed with sterile saline plus providone-iodine (Butler Schein) or chlorhexidine at least three times per week. Treatment with 5-fluorouracil, and occasional dura scrapes, were performed weekly to maintain thin and healthy dura for ease of electrode penetration.

Visual stimulation. Visual stimuli were generated using a VSG/5 system (Cambridge Research Systems). Stimuli were presented on a gamma-calibrated Sony monitor placed 700 mm in front of animals' eyes. The refresh rate of the monitor was 140 Hz and the mean luminance was 38 cd m^{-2} . The monitor was the sole source of illumination in the room containing the animal. All stimuli were presented under binocular viewing conditions.

Behavioural training. Animals were trained to perform fixation and contrast-change detection tasks for juice rewards using standard operant conditioning. Eye position was monitored by an infrared video eye tracker (Applied Science Laboratories) with a refresh rate of 240 Hz. If animals' eye positions deviated by more than 0.35° at any point during a task trial, the trial was aborted. Fixation tasks required animals to maintain central fixation on a dot while drifting gratings were placed within the receptive fields of recorded neurons in order to measure neuronal visual physiology. For fixation tasks, trials were interleaved with a minimum 1-s period during which the monitor was mean luminance grey and animals were allowed to move their eyes freely.

Attention task. Animals were trained to detect a contrast change in one of two gratings and report the change using a joystick. Fixation was maintained throughout the duration of the trial, including the answer window. Animals were instructed to attend either to a drifting sinusoidal grating presented within (attend-towards condition) or an identical grating presented outside (attend-away condition) the receptive fields of the recorded neurons (Fig. 1a). The two gratings were identical and set to the orientation and spatial-frequency preference of the recorded neurons, and were always placed at an equal distance from the central fixation dot. Attention trials were run in blocks of 10 trials of each attention condition (attend towards or attend away) and the colour of the central fixation dot provided a cue to the animals, of where to allocate their attention. Trials progressed as follows. Trials were separated by a minimum of 1,000-ms, during which time the luminance of the monitor was maintained at background levels (mean luminance grey) and the animals were allowed to move their eyes freely, after which the monkeys could initiate a new trial by moving a joystick from the centre position to a side position (left or right of centre). Animals were required to maintain the joystick in the side position throughout the duration of the trial until the answer period; premature joystick movements caused trials to abort. After initial movement of the joystick to the side, a central fixation dot was displayed, to which animals directed their gaze. Five-hundred milliseconds after the onset of central fixation, two gratings appeared on the monitor, one inside and one outside the receptive field of recorded neurons. The two gratings were presented for a variable amount of time, between 1,200 and 2,500 ms, determined on a trial-by-trial basis according to a hazard function, with a mean at 1,700 ms. Following the period of visual stimulation by drifting gratings as determined by the hazard function, one of the two gratings increased in contrast by 10%. Both gratings remained on the monitor during a 500-ms answer window in which animals signalled detection of the contrast change by moving the joystick to the original central position. Only correct detection of the contrast change, indicated by a correct joystick movement, while also maintaining central gaze fixation throughout the answer period, was rewarded with juice. Trained animals typically performed with a success rate of 70% or above, discounting aborted trials. Trials were aborted when animals moved their eyes by more than 0.35° during any part of the trial or made a joystick movement before the contrast change. We found no significant differences in the proportion of aborted trials across attention conditions or across shock and non-shock trials ($P > 0.5$). Importantly, before the contrast change, visual stimulation was equal across attention conditions such that the only variable across conditions was the location to which the animal directed covert spatial attention.

Across blocks of trials, 95% of total trials were cued validly, wherein the contrast change occurred at the attended location signified by the colour of the fixation dot. In the remaining 5% of trials, the fixation dot colour cue was invalid and the contrast change occurred at the unattended location. Reaction times were measured as

the time between the contrast change and movement of the joystick back to the central position. For each animal, reaction times were compared across validly and invalidly cued trials. Reaction time data were computed for all sessions including at least 50 correct trials in each attention condition (attend towards and attend away). Reaction time values were: monkey B, reaction time (valid) = 363 ± 5 ms, reaction time (invalid) = 401 ± 14 ms (monkey B, $P < 0.01$); monkey O, reaction time (valid) = 387 ± 7 ms, reaction time (invalid) = 435 ± 21 ms (monkey O, $P < 0.05$). In addition, both monkeys correctly detected grating contrast changes significantly more often in validly compared to invalidly cued trials ($P < 0.03$).

Electrical stimulation. As described in detail previously^{31,32}, stimulating electrodes were semi-chronically implanted within parafoveal regions of the LGN. Stimulating electrodes were placed in precise retinotopic alignment with recording electrodes in V1, such that receptive fields of neurons at each location were within less than 2° of one another in visual space. Single platinum or iridium stimulating electrodes (FHC) with less than 1 mm of tip exposure were placed within the LGN such that both magnocellular and parvocellular thalamocortical neuronal populations were activated. Placement of stimulating electrodes was guided and verified by recording visual responses from LGN neurons during and after implantation. Stimulation was generated by an isolated pulse stimulator (AM Systems) and included a single, brief (0.2-ms), biphasic current pulse (approximately 10 to 200 μA) delivered once every 5 s during collision testing (described below) and once per shock trial in the attention task (described below).

To locate putative TCR neurons, V1 recording electrodes were advanced slowly while shocks were delivered at regular intervals (every 5 s). TCR neurons were identified by the presence of short-latency (less than 6 ms) feedforward postsynaptic spikes in both collision and non-collision modes of stimulation, as described in detail previously^{31,32}. Shock-evoked postsynaptic spike latencies were calculated as the time between the shock in the LGN and the postsynaptic spike. Magnocellular- and parvocellular-recipient neurons did not differ in their postsynaptic spike latencies (Fig. 2b; $P > 0.5$).

Collision testing was carried out to distinguish whether or not cortical neurons were activated by the arrival of orthodromically or antidromically propagated spikes after electrical stimulation. Collision tests were performed while the animals performed fixation tasks, or when they were free to move their eyes and the monitor was mean luminance grey. Shock current was set so that shocks evoked spikes approximately 35% of the time, on average, regardless of the behavioural condition or stimulus display. It was important to titrate the shock current for each individual TCR neuron such that shock-evoked postsynaptic spikes occurred on a fraction of trials in order to avoid floor or ceiling effects in the attention experiment. In 76 out of 80 recording sessions, shock strength was held constant across collision and attention testing conditions; in four sessions, shock strength was decreased for attention trials relative to collision trials.

Electrical stimulation during the attention task. In 70% of attention trials, a single shock (parameters as above, current set per TCR neuron) was delivered between 1,000 and 1,200 ms after the onset of the two drifting sinusoidal stimuli and before the contrast change. Shocks occurred at the same time in the stimulus cycle for both attend-towards and attend-away trials. The precise timing of the shock was set to match the neuron's peak response to the periodic stimulus. In this way, electrical stimulation occurred while the recorded neuron was excited (rather than suppressed) by the visual stimulus. Shocks were delivered towards the end of the visual stimulation period because animals' reaction times decreased systematically with increasing visual stimulation duration, suggesting that animals exerted greater attention towards the end of each trial (data not shown).

The efficacy of shocks in evoking postsynaptic spikes was determined for each TCR neuron based on all shock trials in the attention task, including attend-towards and attend-away trials (Supplementary Fig. 1a). Magnocellular- and parvocellular-recipient neurons did not differ in their shock-evoked spike efficacies (Supplementary Fig. 1a; $P > 0.05$).

Animals did not make voluntary or involuntary eye movements in response to electrical stimulation of the LGN (which would have resulted in aborted trials, as shocks occurred before grating-contrast changes), and electrical stimulation did not affect performance on the attention task. As electrical stimulation parameters were reduced such that the average efficacy of shock-evoked postsynaptic spikes was approximately 35%, and because the same proportion of trials included shocks in both attention conditions, it is unlikely that shocks induced visual percepts that interfered with animals' behaviour or changes in neuronal responses across attention conditions.

Electrophysiological recordings. Recordings from V1 neurons were made using single platinum-in-glass electrodes (Alpha Omega) or a Mini Matrix multi-electrode array of five platinum-in-quartz electrodes (Thomas Recording). Spiking data were amplified and recorded by a PC equipped with a Power 1401 acquisition system and Spike2 software package (Cambridge Electronic Design). For each recording session, the first step was the identification of putative TCR neurons

(described above). The second step involved characterizing the visual physiology of recorded neurons. This was accomplished by presenting drifting sinusoidal gratings that varied in orientation, contrast, spatial frequency or size within the centre of the receptive field while animals performed the fixation task. Gratings were presented for 1 to 2 s per trial, and trials were repeated at least two times. To generate response functions for orientation (0 to 360°), contrast (0 to 100%), spatial frequency (0.2 to 4 cycles per degree), and size (0.2 to 10°), individual parameters were increased in 10- to 15-step increments while all other parameters were held constant. Once optimal stimulus parameters were determined, optimal gratings were presented for 2 s per trial, to determine the precise time (the time of peak response to the periodic stimulus) to deliver electrical stimulation (see above). Finally, neurons were recorded while animals performed the attention task. Gratings drifted at 4 Hz and were of optimal orientation, spatial frequency and approximately two to four times the size of the receptive field to accommodate small shifts in eye position (less than 0.35°). Grating contrast was 70% for putative parvocellular-recipient neurons and between 10 and 25% for putative magnocellular-recipient neurons. For recording sessions in which multiple neurons were recorded simultaneously using the multi-electrode array, gratings parameters were set to stimulate the maximum number of cells as optimally as possible and grating sizes were set to cover the receptive fields of all recorded neurons (never greater than 2°, as receptive-field locations always overlapped greatly). If putative magnocellular- and parvocellular-recipient neurons were recorded simultaneously, grating contrast was set to an intermediate value of approximately 40 to 50%. For the contrast-change detection portion of the attention task, contrast always increased by 10% of the starting contrast.

Data analyses. All recorded spikes were sorted offline using principal components analysis (Spike2 software standard algorithms). Recordings were made from 161 neurons in V1, of which 61 neurons were identified as TCR neurons based on responses during collision testing. Out of these 61 TCR neurons, 22 were recorded with single electrodes (15 from monkey B, 7 from monkey O), and 39 were recorded with the multi-electrode array (30 from monkey B, 9 from monkey O). Twenty-nine additional neurons were classified as putative TCR neurons in multi-electrode recordings (21 from monkey B, 8 from monkey O) based on shock-evoked responses during the attention task. For these neurons, shock-evoked postsynaptic responses were not consistent during collision testing because the shock current was set to accommodate the activity of a different, identified neighbouring TCR in the same recording session. However, during the attention task, shocks systematically evoked postsynaptic spikes at fixed latencies consistent with monosynaptic responses. As TCR neurons and putative TCR neurons did not differ significantly from each other in their percentages of synchronous evoked spikes in attend-towards or attend-away conditions ($P > 0.2$), both groups of neurons were included in subsequent analyses of synchronized and correlated spiking across attention conditions (see below). Receptive fields for all recorded TCR neurons were located in the lower left visual hemifield at parafoveal eccentricities. There were no differences in physiological response properties or attentional modulation of neurons recorded in the two monkeys, and thus, neurophysiological data from both monkeys were combined for all analyses.

TCR neurons were designated as magnocellular- or parvocellular-recipient neurons by their C_{50} values (contrast that evoked a half-maximum response) measured from exponential fits to their contrast-response functions. Magnocellular-recipient neurons ($n = 36$) were classified as neurons with C_{50} values of less than 30% and parvocellular-recipient neurons ($n = 25$) were classified by C_{50} values of greater than 30%. All neurons were classified as simple or complex cells based on the ratio of the first Fourier coefficient (f_1) to mean (f_0) response, where simple cells have $f_1:f_0 > 1$ and complex cells have an $f_1:f_0 < 1$ (ref. 33). Mean $f_1:f_0$ ratios were: magnocellular-recipient neurons, 0.4 ± 0.04 ; parvocellular-recipient neurons, 1.2 ± 0.06 ($P < 0.001$). Subsequent analyses of firing rates for visual physiological characterizations and attention-index calculations were performed on the f_1 (simple cells) or mean (complex cells) response of each neuron. Orientation-tuning bandwidth was determined by calculating the peak half-width at half-height of Gaussian fits to individual orientation-tuning curves.

Firing rates were measured during inter-trial intervals and during fixation before the presentation of gratings during the attention task. There were no significant differences between magnocellular and parvocellular-recipient neurons or across attention conditions for these firing-rate measurements ($P > 0.7$). Firing rates were also measured during visual-stimulus presentation before the contrast change during attention trials. Magnocellular-recipient neurons had significantly higher firing rates during this period ($P < 0.05$; mean = 246 ± 19 spikes per s) compared to parvocellular-recipient neurons (mean = 185 ± 23 spikes per s), but firing rates assessed over this period for both magnocellular- and parvocellular-recipient neurons did not differ between attend-towards and attend-away conditions ($P > 0.75$).

For all analyses involving an examination of changes in the percentage of shock-evoked spikes, trials were sorted according to whether or not a spike occurred at the specific and fixed postsynaptic spike latency for each individual TCR neuron. Shock-evoked postsynaptic spikes were determined using a 2-ms window aligned by the spike latency for the neuron under study. The proportions of shocks that evoked a postsynaptic spike in each attention condition were determined for each TCR neuron. We also calculated the proportion of non-shock trials in which a spike occurred at the same latency for each attention condition to allow for a comparison between the number of spikes within the latency window with and without electrical stimulation (Supplementary Fig. 1b). As expected, shocks elicited significantly more spikes at the postsynaptic response latency compared to the number of spikes that occurred in the same time window during non-shock trials for the attend-towards condition ($P < 0.01$). However, there were no differences in the number of spikes that occurred at the postsynaptic response latency between non-shock and shock trials for the attend-away condition.

For each TCR neuron, an attention-index value was calculated as the difference (attend towards – attend away) divided by the sum (attend towards + attend away) of average spiking activity over a specified duration of visual stimulation and always before the earliest opportunity for contrast change in attention trials. Importantly, attention-index value calculations always included the time windows corresponding to the shock and shock-evoked postsynaptic responses (which occurred between 1,000 and 1,200 ms in all trials). We calculated attention-index values over long and short durations of visual stimulation: 0 to 1,200 ms, 600 to 1,200 ms, 850 to 1,200 ms, and 1,000 to 1,200 ms after the onset of grating stimulation. When we calculated attention-index values based on the firing rate over long durations (0 to 1,200 ms and 600 to 1,200 ms after grating presentation), we observed no changes in attention index across attention conditions (Supplementary Fig. 1c). When we calculated attention index over short durations (850 to 1,200 ms and 1,000 to 1,200 ms), we observed very small shifts in attention index for magnocellular-recipient neurons only. We compared attention-index values for short durations to spike efficacy values and observed no relationship between overall changes in firing rate with attention and changes in spiking efficacy with attention (Supplementary Fig. 1d).

To examine the influence of attention on the temporal precision of thalamo-cortical communication across our sample of recorded TCR neurons, we aligned the ongoing spiking responses of each TCR before and after individual shocks, so that time = 0 corresponded to the time when the TCR neuron was expected to produce a shock-evoked postsynaptic response (determined from the shock-evoked postsynaptic latency). This allowed for averaging of spiking activity (or differential spiking activity: activity in attend-towards trials minus activity in attend-away trials) across magnocellular- and parvocellular-recipient neurons with different feedforward spike latencies. To assess the jitter in postsynaptic spike timing across attention conditions, we also plotted time courses surrounding shocks separately for each attention condition (Supplementary Fig. 2a). To demonstrate differential spiking activity for magnocellular- and parvocellular-recipient neurons, we reported two times the standard deviation of average spiking activity before the shock. In all cases, error ranges were reported as standard errors of the mean. We also separated magnocellular- and parvocellular-recipient neurons into two subpopulations based on the presence of a negative dip in spiking activity just before the postsynaptic spike (15 magnocellular-recipient and 10 parvocellular-recipient neurons displayed negative dips). TCR neurons were classified as 'dip' neurons when differential spike count values at –2 or –1 ms time points (relative to the time of the shock-evoked monosynaptic spike) were less than two times the standard deviation of the mean activity before the shock. There was no relationship between spike latency and whether or not a neuron displayed a dip ($P > 0.75$). Moreover, across the sample of dip neurons, there was a range of spike latencies, including latencies longer than the period of the dip, indicating that dips were not systematically a consequence of the shock-induced stimulus artefact obscuring our ability to detect spikes. Supplementary Fig. 2a, b illustrates spiking activity surrounding the postsynaptic spike for dip and no-dip TCR subpopulations (with magnocellular- and parvocellular-recipient groups plotted separately in Supplementary Fig. 2b). We fit Gaussian equations to positive peaks (corresponding to shock-evoked postsynaptic spikes) for dip and no-dip population average curves in order to calculate the width at half-height values for each fit.

During 18 sessions (13 from monkey B, 5 from monkey O) we used a five-channel multi-electrode array with independently movable microelectrodes (Thomas Recording Mini-Matrix system) and recorded from 39 TCR neurons and 29 putative TCR neurons (see above). In 3 of the 18 sessions we recorded from a single TCR neuron. Sessions with paired recordings were as follows: 3 sessions with 1 pair, 1 session with 2 pairs, 2 sessions with 3 pairs, 1 session with 4 pairs, 1 session with 5 pairs, 2 sessions with 6 pairs, 3 sessions with 7 pairs, and 2 sessions with 9 pairs (total = 71 pairs across 15 sessions; mean = 3.3 cells recorded per session). Out of the total 71 pairs, 25 received common input (that is, cross-correlograms contained a central

narrow peak at time = 0) and 46 received independent input. We examined all 50 of the possible common input pairings (in both directions) and used 45 pairings for the analysis of attentional modulation of correlated inputs for pairs receiving common presynaptic input (5 pairings were excluded for lack of sufficient spikes and/or noisy correlograms). To determine whether recording sessions with several pairs did not systematically bias the results, we compared attentional modulation of correlated spikes across recording sessions with greater than 2 pairings and found no differences in attentional modulation of correlated spikes across sessions ($P > 0.05$). We recorded a total of 48 magnocellular–magnocellular pairs (21 received common input, 27 received independent input); a total of 11 parvocellular–parvocellular pairs (2 received common input, 9 received independent input); and a total of 12 magnocellular–parvocellular pairs (2 received common input, 10 received independent input).

We calculated the probability that shocks would evoke synchronous spikes in both TCR neurons across attention conditions. Importantly, our criteria for defining synchronous spikes were strict. Trials with synchronous spikes were those where each TCR neuron fired a shock-evoked postsynaptic spike at its identified postsynaptic spike latency ($n = 71$ pairs).

Cross-correlation analysis was used to determine whether pairs of TCR neurons received input from independent (that is, separate) LGN axons or input from common LGN axons with presumed anatomical divergence. Cross correlations were calculated on a trial-by-trial basis for each of the 71 TCR pairings using spiking data from the 600- to 1,200-ms period of visual stimulation. Shuffled cross correlations were also calculated for each pairing by correlating spikes from neuron A shifted by one stimulus cycle (250 ms) compared to those of neuron B. For each trial, the shuffled correlogram was subtracted from the original correlogram. By employing trial-specific shuffle corrections, we eliminated spike correlations emerging from slow co-variations in neuronal firing rate³⁴ and co-activation from a common visual stimulus (stimulus-dependent correlations). Shuffle-subtracted correlograms were reported as average percentages of total spikes for each attention condition.

Out of 71 pairs of simultaneously recorded TCR neurons, 25 pairs (17 from monkey B, 8 from monkey O) displayed narrow peaks (< 3 ms) in their cross correlograms, centred at time zero, indicating the pair received common presynaptic input, presumably from feedforward axons with branches that contacted both neurons. In addition to magnocellular–magnocellular ($n = 21$) and parvocellular–parvocellular pairs ($n = 2$), we also encountered mixed magnocellular–parvocellular pairs ($n = 2$) receiving common input. Neurons in mixed magnocellular–parvocellular-recipient pairs were located within close proximity to one another in cortical depth (within $75 \mu\text{m}$), and tended to display similar contrast sensitivity and/or orientation selectivity, suggesting that neurons in these mixed pairs were both located near the laminar border between layers $4C\alpha$ and $4C\beta$.

For the analysis of attentional modulation of TCR pairs receiving common input, we calculated the difference in peak area, measured in 3 bins of 1-ms width centred at time = 0, between shuffle-corrected cross-correlograms in each attention condition. Cross-correlogram peak heights in each attentional condition corresponded to roughly 3% of total spikes (Fig. 4c) and the average attention-mediated reduction in peak height across all cells was $-0.3 \pm 0.2\%$ suggesting that attention caused a 10% reduction in correlated spiking resulting from divergent input from a common presynaptic source.

To compare actual measured percentages of synchronous shock-evoked postsynaptic spikes to predicted percentages, we compared the measured incidences of synchronous spikes (as described above) with the product of each individual neuron's probability of firing a postsynaptic spike in response to the shock (for example, Fig. 2d). We then calculated the difference (actual – predicted) in the occurrence of synchronous spikes for common input TCR pairs and independent-input TCR pairs across attention conditions.

Approximation of overall attentional modulation of signal-to-noise ratio was calculated as: (change in percentage signal)/(change in percentage noise), which translates to $(1 + \text{average percentage increase in shock-evoked synchronized spikes}) / (1 - \text{average percentage decrease in synchronized spikes in cross-correlogram peaks})$. The average percentage increase in shock-evoked synchronized spikes = 8%, and average percentage decrease in synchronized spikes in cross-correlogram peaks = 10% for the population of recorded TCR pairs (see above).

Statistics. Parametric or non-parametric comparisons tests (t -test or rank-sum test, respectively) were used for all two-sample comparisons depending on the distribution normality of the samples tested. To test for distribution normality, one-sample Kolmogorov–Smirnov tests were used. To examine whether any given distribution of data differed from an equivalent normal distribution, the sample distribution was compared to a normal distribution with the same standard deviation as the sample. Accordingly, the Kolmogorov–Smirnov test compared a sample distribution to a proposed continuous distribution defined by the same range and variance parameters as the sample.

31. Briggs, F. & Usrey, W. M. A fast, reciprocal pathway between the lateral geniculate nucleus and visual cortex in the macaque monkey. *J. Neurosci.* **27**, 5431–5436 (2007).
32. Briggs, F. & Usrey, W. M. Parallel processing in the corticogeniculate pathway of the macaque monkey. *Neuron* **62**, 135–146 (2009).
33. Skottun, B. C. *et al.* Classifying simple and complex cells on the basis of response modulation. *Vision Res.* **31**, 1078–1086 (1991).
34. Brody, C. D. Slow covariations in neuronal resting potentials can lead to artefactual fast cross-correlations in their spike trains. *J. Neurophysiol.* **80**, 3345–3351 (1998).

High frequency of functional extinctions in ecological networks

Torbjörn Säterberg¹, Stefan Sellman¹ & Bo Ebenman¹

Intensified exploitation of natural populations and habitats has led to increased mortality rates and decreased abundances of many species^{1,2}. There is a growing concern that this might cause critical abundance thresholds of species to be crossed^{1,3–5}, with extinction cascades and state shifts in ecosystems as a consequence^{4,6,7}. When increased mortality rate and decreased abundance of a given species lead to extinction of other species, this species can be characterized as functionally extinct even though it still exists. Although such functional extinctions have been observed in some ecosystems^{3,4,8}, their frequency is largely unknown. Here we use a new modelling approach to explore the frequency and pattern of functional extinctions in ecological networks. Specifically, we analytically derive critical abundance thresholds of species by increasing their mortality rates until an extinction occurs in the network. Applying this approach on natural and theoretical food webs, we show that the species most likely to go extinct first is not the one whose mortality rate is increased but instead another species. Indeed, up to 80% of all first extinctions are of another species, suggesting that a species' ecological functionality is often lost before its own existence is threatened. Furthermore, we find that large-bodied species at the top of the food chains can only be exposed to small increases in mortality rate and small decreases in abundance before going functionally extinct compared to small-bodied species lower in the food chains. These results illustrate the potential importance of functional extinctions in ecological networks and lend strong support to arguments advocating a more community-oriented approach in conservation biology, with target levels for populations based on ecological functionality rather than on mere persistence^{8–11}.

Recent research points to the importance of maintaining populations at ecologically effective sizes, that is, population sizes large enough to maintain critical interactions and hence ensure the functionality and stability of ecological communities^{5,8–12}. An ecologically effective population size (EEP) of a species can be defined as the population size below which one or more other species in the ecological community go extinct. Thus, a species can be said to go ecologically or functionally extinct when its population has fallen below this size. More precisely, we here define functional extinction in the following way: a focal species goes functionally extinct when a reduction in its population size—caused by an increased mortality rate—leads to the extinction of another species in the community (see Supplementary Fig. 1 for an illustrative example). The strong evidence of trophic cascades and regime shifts in a wide variety of ecosystems following the reduction in population size of a species^{3,4,7,13–15} clearly indicate that functional population size thresholds are present in many ecosystems. However, which species are most likely to go functionally extinct and if density thresholds differ among species depending on their traits remains unexplored. Here we present an analytical method of deriving EEPs for species in ecological communities. The EEP of a species is found by increasing its mortality rate until an extinction occurs in the community (see Methods).

We apply this analytical method to derive EEPs of species in eight natural food webs (Supplementary Table 2) and in sequentially

assembled model food webs. Model webs were generated by relating parameter values to the average body mass of species and letting dynamical constraints act during the web build-up (Supplementary Methods, Supplementary Table 1 and Supplementary Fig. 9). Using this approach, the binary network of who eats whom (the network topology) is an outcome of the dynamical assembly process rather than being pre-set before adding dynamics. The topology as well as the mass-abundance slope (log abundance as a function of log average body mass) in these model food webs is within the ranges observed for real food webs (Supplementary Figs 2 and 3). Models of natural food webs were parameterised using allometric relationships and, where available, data from gut analysis¹⁶ (Supplementary Methods). For each food web we record the identity of the focal species (the species whose mortality rate is increased) and the extinct species at the analytically calculated functional extinction thresholds (EEP). These data are used to investigate (1) the topological relationship between the focal species and the extinct species; (2) how different species characteristics affect the probability that a species goes functionally rather than numerically extinct (that is, the probability that increased mortality rate of a species leads to extinction of another species rather than of the species itself); (3) what determines a species ecologically functional population size.

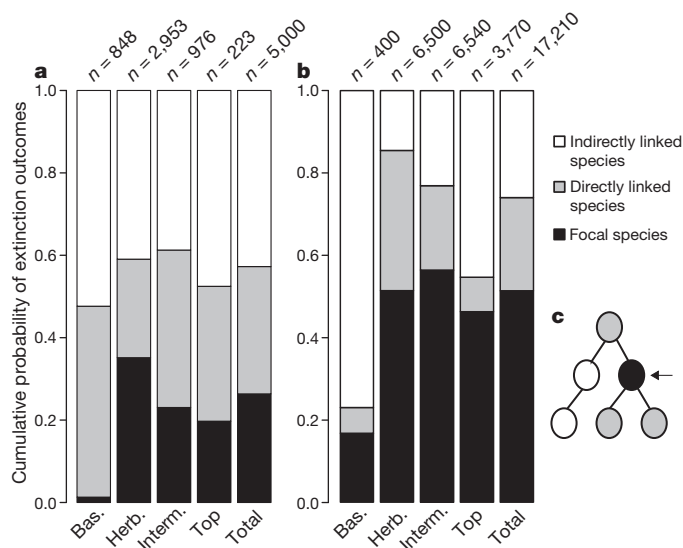


Figure 1 | Increased mortality rate of a species most often leads to extinction of another species. a, b, Proportion of extinctions affecting focal species (black), directly (grey) and indirectly (white) linked species in theoretical food webs (a) and natural food webs (b). Focal species belong to one of the following trophic groups: basal species (primary producers); herbivores (species feeding only on primary producers); intermediate species (consumers that are neither top nor herbivore species) or top species (species that are not consumed). c, Picture illustrating the extinction outcome categories in a and b. Sample sizes are given above columns.

¹Linköping University, Department of Physics, Chemistry and Biology, Division of Theoretical Biology, SE-58183 Linköping, Sweden.

We find that there is a high probability—on average 0.49 for the natural food webs and 0.72 for the model webs—that the species to be the first to go extinct is not the one whose mortality rate is increased but instead some other species in the food web (Fig. 1, Supplementary Figs 1 and 4). There is variation in the frequency of functional extinctions across the natural food webs (31.8%–80.3%), but no clear effects of habitat type (Supplementary Fig. 4). High frequency of functional extinctions also occurs in model food webs with nonlinear functional responses of consumers, demonstrating the robustness of the pattern (Supplementary Fig. 5). Furthermore, functional extinctions occur at all trophic levels (Fig. 1, Supplementary Figs 1, 4 and 7), demonstrating that both top-down and bottom-up processes are important. A relatively high proportion of all first extinctions are of species that are not directly linked to the species whose mortality rate was increased and population density decreased (Fig. 1), that is, the extinct species is neither a consumer of nor consumed by the focal species. Recent theoretical studies suggest that the effect of extinction of a given species on another species depends on the trophic distance between them^{17,18}. In highly connected ecological networks, species going secondarily extinct were often more than one link away from the initially lost species¹⁸, consistent with our results.

Critical density thresholds of species, set by direct and indirect interactions among species, have been documented in real food webs. For instance, a recent study found strong evidence of an ecologically effective population size for sea otters, below which sea urchins burst in density due to released predation pressure from the otters, which in turn led to overexploitation of kelp by the sea urchins and accompanying impoverishment and collapse of the kelp-forest ecosystem⁸. Examples like the kelp-forest food web have led several researchers to argue that target population sizes should be set based on ecological criteria and not merely on demographic criteria^{1,9,10}. In light of this, it might be interesting to set our functional extinction thresholds in relation to the threat categories in the International Union for Conservation of Nature (IUCN) red list for threatened species¹⁹. For instance, a density reduction of over 30% over a period of three generations would categorize a species as ‘Vulnerable’ according to IUCN definitions. We find that, on average, 26.1% of the species in the

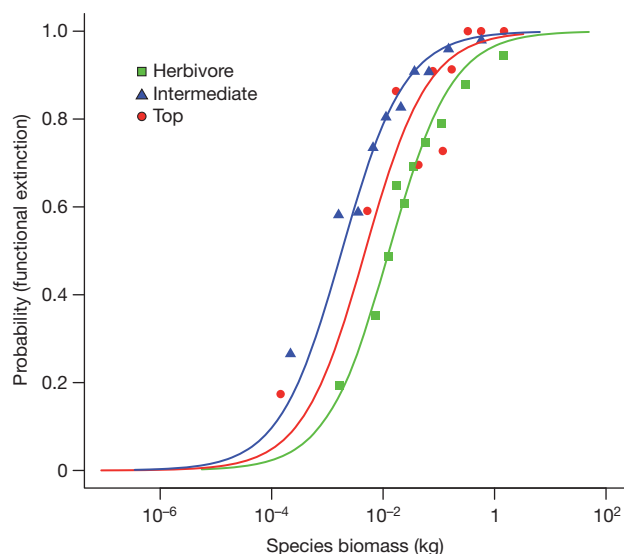


Figure 2 | Functional extinction is more likely than numerical extinction in species with large biomass. The probability that the extinction of a species is functional (rather than numerical) plotted against the biomass of the species, for species in different trophic groups. Symbols represent means for binned data and lines represent model predictions based on a logistic mixed effect model (Supplementary Methods and Supplementary Table 6). Each bin contains 10% of the observations ranked according to species biomass, for each trophic group. Sample size $n = 5,000$.

natural webs and 54.0% of the species in the model food webs go functionally extinct even before losing 30% of their individuals, thus indicating that species might often go functionally extinct long before being recognized as threatened.

To identify which species are more likely to go functionally than numerically extinct, a large number of species traits (Supplementary Tables 3, 4 and 5) were analysed. Of these traits species equilibrium biomass and trophic level (trophic group) were identified as major predictors of functional extinction in a classification tree analysis of the theoretical webs (Supplementary Fig. 6). We also fitted a logistic mixed effect model with species equilibrium biomass and trophic group as predictors (see Supplementary Methods). Both analyses show that the probability that a species will go functionally rather than numerically extinct is positively related to the equilibrium biomass of the species, with an additional effect of its trophic group belonging (Fig. 2, Supplementary Fig. 6 and Supplementary Table 6). A positive relationship between the biomass of a species and the probability of functional rather than numerical extinction is also present in the natural food webs (except for the Celtic Sea food web where no significant relationship was found, Supplementary Table 7).

We find that the reduction in abundance of a species needed to cause extinction of another species is strongly and inversely related

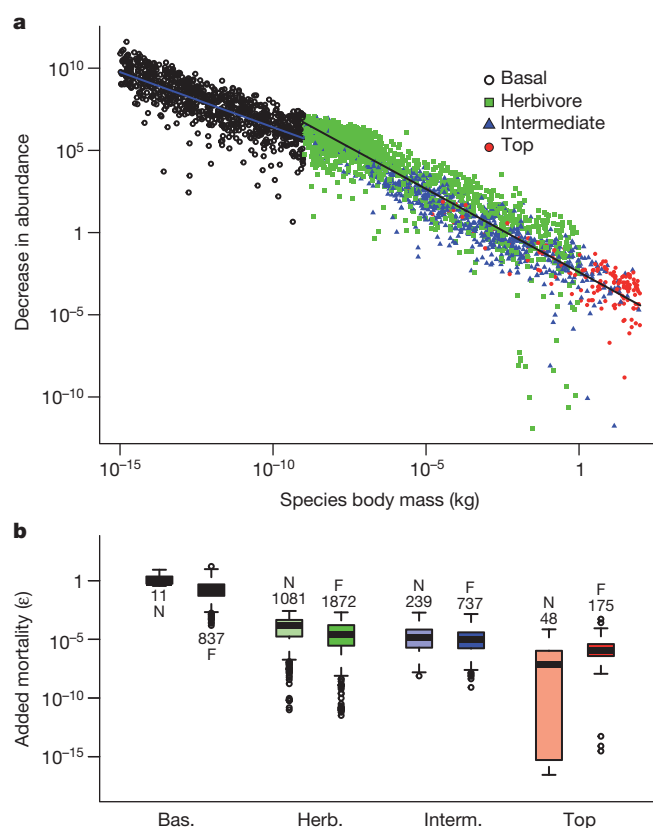


Figure 3 | Species at the top of food webs can only lose few individuals and sustain small extra mortality rates before going functionally extinct.

a, Decrease in abundance of a species needed to cause extinction of another species in the web as a function of the body mass of the species. Blue and black lines are predictions based on linear mixed effect models for basal ($n = 837$) and consumer ($n = 2,670$) species, respectively (Supplementary Methods and Supplementary Table 8). **b**, Boxplots showing the amount (ϵ) by which mortality rate was increased in species that went numerically (N) and functionally (F) extinct, respectively, for species in different trophic groups. Bottom and top of boxes are the first and third quartiles (Q1 and Q3), respectively, and the line within boxes is the median (second quartile, Q2); lower and upper whiskers are $\max(Q1 - 1.5 \times IQR, \min(\epsilon))$ and $\min(Q3 + 1.5 \times IQR, \max(\epsilon))$, respectively (IQR is interquartile range). Sample sizes are given adjacent to boxplots.

to its body mass (Fig. 3a and Supplementary Table 8). Turning to mortality, we find that the increase in mortality rate of a species needed to cause extinction of another species in the food web decreases with increasing trophic height and body mass of the species (Fig. 3b and Supplementary Fig. 8). Thus, for large-bodied species even a comparably small increase in their mortality rate can lead to the extinction of another species in the food web. Furthermore, comparing functionally and numerically extinct species, we find that the increase in mortality rate causing functional extinction of species is approximately the same as that causing numerical extinction of species (Fig. 3b and Supplementary Fig. 8).

We have here focused on deriving functional extinction thresholds of species. Once the functional extinction threshold is crossed and a species is lost from the food web, the system might lose its stability. In the worst case the loss of the first species might trigger an extinction cascade^{3,18,20–24}. Preliminary results indicate that this risk might be substantial: we find that 53.5% of all functional extinctions have led to model food webs that are locally unstable. For the natural webs the corresponding figure is, on average, 53.7%. This indicates that relatively small changes in the mortality rates of species can have even larger effects than what we report here. To conclude, our study demonstrates the potential importance of functional extinctions in ecological networks and lends strong support to arguments put forward for a more system-oriented approach when setting target levels for populations of species in various ecosystems^{9–11}.

METHODS SUMMARY

We generated feasible and locally stable model food webs, each with 50 interacting species, using a sequential assembly approach²⁵ (Supplementary Methods, Supplementary Table 1 and Supplementary Fig. 9). Eight natural food webs are also included in the analysis (Supplementary Table 2). Food web dynamics are described by a generalized Lotka-Volterra model and model parameters are derived from body sizes of species using allometric relationships and, where available, data from gut analysis¹⁶ (see also Supplementary Methods).

To identify which species are more likely to go functionally than numerically extinct following an increased mortality rate, we use a classification tree analysis and a generalized linear mixed effect model (GLMM). The classification tree analysis²⁶ is used to identify the most important predictors (species traits) among a large number of predictors (described in Supplementary Tables 2, 3 & 4). A GLMM²⁷ is then used for a detailed statistical analysis of the most important predictors where we account for the clustered nature of data (clustered within food webs).

EEPs were analytically derived for each species within each food web (see Methods section). We investigate patterns of EEPs using data for species in which added mortality leads to extinction of another species in the food web rather than an extinction of the species itself. The number of individuals that can be lost in a species before it goes functionally extinct is analysed using a linear mixed effect model (LMM) with species body mass as covariate.

Full Methods and any associated references are available in the online version of the paper.

Received 14 December 2012; accepted 9 May 2013.

Published online 7 July 2013.

1. Jackson, J. B. C. *et al.* Historical overfishing and the recent collapse of coastal ecosystems. *Science* **293**, 629–637 (2001).
2. Myers, R. A. & Worm, B. Rapid worldwide depletion of predatory fish communities. *Nature* **423**, 280–283 (2003).

3. Terborgh, J. & Estes, J. A. (eds) *Trophic Cascades: Predators, Prey, and the Changing Dynamics of Nature* (Island Press, 2010).
4. Estes, J. A. *et al.* Trophic downgrading of planet earth. *Science* **333**, 301–306 (2011).
5. Anderson, S. H., Kelly, D., Ladley, J. J., Molloy, S. & Terry, J. Cascading effects of bird functional extinction reduce pollination and plant density. *Science* **331**, 1068–1071 (2011).
6. Barnosky, A. D. *et al.* Approaching a state shift in Earth's biosphere. *Nature* **486**, 52–58 (2012).
7. Scheffer, M., Carpenter, S., Foley, J. A., Folke, C. & Walker, B. Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001).
8. Estes, J. A., Tinker, M. T. & Bodkin, J. L. Using ecological function to develop recovery criteria for depleted species: sea otters and kelp forests in the aleutian archipelago. *Conserv. Biol.* **24**, 852–860 (2010).
9. Soulé, M. E., Estes, J. A., Berger, J. & Del Rio, C. M. Ecological effectiveness: conservation goals for interactive species. *Conserv. Biol.* **17**, 1238–1250 (2003).
10. Soulé, M. E., Estes, J. A., Miller, B. & Honnold, D. L. Strongly interacting species: conservation policy, management, and ethics. *Bioscience* **55**, 168–176 (2005).
11. Sabo, J. L. Population viability and species interactions: life outside the single-species vacuum. *Biol. Conserv.* **141**, 276–286 (2008).
12. McConkey, K. R. & Drake, D. R. Flying foxes cease to function as seed dispersers long before they become rare. *Ecology* **87**, 271–276 (2006).
13. Casini, M. *et al.* Trophic cascades promote threshold-like shifts in pelagic marine ecosystems. *Proc. Natl Acad. Sci. USA* **106**, 197–202 (2009).
14. Frank, K. T., Petrie, B., Choi, J. S. & Leggett, W. C. Trophic cascades in a formerly cod-dominated ecosystem. *Science* **308**, 1621–1623 (2005).
15. Smith, A. D. *et al.* Impacts of fishing low-trophic level species on marine ecosystems. *Science* **333**, 1147–1150 (2011).
16. Berg, S., Christianou, M., Jonsson, T. & Ebenman, B. Using sensitivity analysis to identify keystone species and keystone links in size-based food webs. *Oikos* **120**, 510–519 (2011).
17. Berlow, E. L. *et al.* Simple prediction of interaction strengths in complex food webs. *Proc. Natl Acad. Sci. USA* **106**, 187–191 (2009).
18. Stouffer, D. B. & Bascompte, J. Compartmentalization increases food-web persistence. *Proc. Natl Acad. Sci. USA* **108**, 3648–3652 (2011).
19. IUCN Red List Categories and Criteria: Version 3.1 (2001), <http://www.iucnredlist.org/technical-documents/categories-and-criteria/2001-categories-criteria>.
20. Paine, R. T. Food web complexity and species diversity. *Am. Nat.* **100**, 65–75 (1966).
21. Borrvall, C., Ebenman, B. & Tomas Jonsson, T. J. Biodiversity lessens the risk of cascading extinction in model food webs. *Ecol. Lett.* **3**, 131–136 (2000).
22. Ebenman, B. & Jonsson, T. Using community viability analysis to identify fragile systems and keystone species. *Trends Ecol. Evol.* **20**, 568–575 (2005).
23. Dunne, J. A., Williams, R. J. & Martinez, N. D. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecol. Lett.* **5**, 558–567 (2002).
24. Colwell, R. K., Dunn, R. R. & Harris, N. C. Coextinction and persistence of dependent species in a changing world. *Annu. Rev. Ecol. Syst.* **43**, 183–203 (2012).
25. Lewis, H. M. & Law, R. Effects of dynamics on ecological networks. *J. Theor. Biol.* **247**, 64–76 (2007).
26. De'ath, G. & Fabricius, K. E. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology* **81**, 3178–3192 (2000).
27. Bolker, B. M. *et al.* Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.* **24**, 127–135 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank N. Virgo, A. Clark, R. Law, O. Petchey, P. Münger, D. Gilljam, A. Curtsdotter and U. Wennergren for comments and discussion. D. Gilljam and A. Curtsdotter also provided computer code for part of the analysis. This project was supported by a Faculty grant from Linköping University to B.E.

Author Contributions T.S. and B.E. designed the project, T.S., B.E. and S.S. performed research, T.S. analysed the data, T.S. wrote the Supplementary Information, T.S. and B.E. wrote the manuscript with contributions from S.S.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.E. (boebe@ifm.liu.se).

METHODS

Study systems—natural and theoretical food webs. We generated 100 feasible and locally stable model food webs, each with 50 interacting species, using a sequential assembly approach²⁵ (Supplementary Methods, Supplementary Table 1 and Supplementary Fig. 9). Eight natural food webs are also included in the analysis (Supplementary Table 2). Food web dynamics are described by a generalized Lotka-Volterra model and model parameters are derived from body sizes of species using allometric relationships and, if available, data from gut analysis¹⁶ (see also Supplementary Methods).

Analytical derivation of ecologically effective population sizes (EEP). Consider an ecological network with S interacting species whose dynamics are described by a generalized Lotka-Volterra model^{28,29}:

$$\frac{d\mathbf{N}}{dt} = \mathbf{N}(\mathbf{r} + \mathbf{A}\mathbf{N}) \quad (1)$$

where $d\mathbf{N}/dt$ is a vector consisting of the rate of change in the density of each species, \mathbf{N} is a column vector consisting of the density of each species (N_i is the density of species i), \mathbf{r} is a vector consisting of the intrinsic growth rate of each of the species and \mathbf{A} is the interaction matrix whose elements are the per capita interaction strengths among species in the network. The interior equilibrium point of the system, $\hat{\mathbf{N}}$, is given by:

$$\hat{\mathbf{N}} = -\mathbf{A}^{-1}\mathbf{r} \quad (2)$$

where \mathbf{A}^{-1} is the inverse interaction matrix.

Now, because of interactions among species, an increase in the mortality rate of one species could have an effect on more than just the species itself^{16,30–32}. To investigate these effects we increase the mortality rate of a species up to the point where an extinction occurs in the ecological network. To increase the mortality rate, we simply decrease the intrinsic growth rate. For each species j , we take the new intrinsic growth rate as $r'_j = r_j - \varepsilon_j$ where ε_j is the smallest amount by which the intrinsic growth rate can be decreased before some species in the network becomes extinct. The new equilibrium densities of the species are then given by:

$$\hat{\mathbf{N}}' = -\mathbf{A}^{-1}(\mathbf{r} - \varepsilon_j \mathbf{e}_j) \quad (3)$$

where \mathbf{e}_j is a column vector with a 1 in the j th row and zero elsewhere. The new equilibrium density of species i (after an increase in the mortality rate of species j) can then be written as:

$$\hat{N}'_i = \hat{N}_i + \varepsilon_j \gamma_{ij} \quad (4)$$

where γ_{ij} is the (i,j) th element of \mathbf{A}^{-1} . Setting $\hat{N}'_i = 0$ in equation (4) gives $\varepsilon_j(i) = -\hat{N}_i/\gamma_{ij}$. We wish to find the smallest positive ε_j for each species j . To do this we simply calculate $\varepsilon_j(i)$ for all species i and find the minimum positive value:

$$\varepsilon_j = \min_i(\varepsilon_j(i) : \varepsilon_j(i) > 0) \quad (5)$$

This value is the minimum amount by which we can increase the mortality rate of species j before some species in the network becomes extinct. Following an increase in the mortality rate of species j , we can calculate a new population size for species j . This new population size can be thought of as a measure of the EEP of the species, as it is the abundance of species j at the point where an extinction occurs in the network due to a change in its mortality rate:

$$\hat{N}_j(\text{EEP}) = \hat{N}_j + \varepsilon_j \gamma_{jj} \quad (6)$$

where γ_{jj} is the (j,j) th element of \mathbf{A}^{-1} .

Statistical analysis. To identify which species are more likely to go functionally than numerically extinct when exposed to an increased mortality rate, we use a classification tree analysis and a generalized linear mixed effect model (GLMM). The classification tree analysis²⁶ is used to identify the most important predictors (species traits) among a large number of predictors (described in Supplementary Tables 2, 3 and 4). A GLMM²⁷ was then used for a detailed statistical analysis of the most important predictors where we account for the clustered nature of data (clustered within food webs).

We investigate patterns in ecologically effective population sizes of different species using data for species where increased mortality leads to extinction of another species in the food web rather than an extinction of the species itself. In a few cases (3.0% of all cases in theoretical food webs and, on average, 2.0% of all cases in the natural food webs) increased mortality rate of a species leads to increased abundance of the species itself—these cases are not included in the analysis. The number of individuals that can be lost in a species before it goes functionally extinct is analysed using a linear mixed effect model (LMM) with species body mass as covariate.

28. Chesson, P. & Kuang, J. J. The interaction between predation and competition. *Nature* **456**, 235–238 (2008).
29. Sahasrabudhe, S. & Motter, A. E. Rescuing ecosystems from extinction cascades through compensatory perturbations. *Nat. Commun.* **2**, 170 (2011).
30. Ives, A. R. & Cardinale, B. J. Food-web interactions govern the resistance of communities after non-random extinctions. *Nature* **429**, 174–177 (2004).
31. Bender, E. A., Case, T. J. & Gilpin, M. E. Perturbation experiments in community ecology: theory and practice. *Ecology* **65**, 1–13 (1984).
32. Montoya, J. M., Woodward, G., Emmerson, M. C. & Sole, R. V. Press perturbations and indirect effects in real food webs. *Ecology* **90**, 2426–2433 (2009).

CORRIGENDUM

doi:10.1038/nature12288

Corrigendum: Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression

Arun Sreekumar, Laila M. Poisson, Thekkelnaycke M. Rajendiran, Amjad P. Khan, Qi Cao, Jindan Yu, Bharathi Laxman, Rohit Mehra, Robert J. Lonigro, Yong Li, Mukesh K. Nyati, Aarif Ahsan, Shanker Kalyana-Sundaram, Bo Han, Xuhong Cao, Jaeman Byun, Gilbert S. Omenn, Debashis Ghosh, Subramaniam Pennathur, Danny C. Alexander, Alvin Berger, Jeffrey R. Shuster, John T. Wei, Sooryanarayana Varambally, Christopher Beecher & Arul M. Chinnaiyan

Nature **457**, 910–914 (2009); doi:10.1038/nature07762

In Fig. 4b of this Article, a typographical error was made in reporting sarcosine levels in the DU145 cell line represented. The y axis values should be in the scale of 0–50 pmoles per 10^6 cells, rather than 0–500 pmoles per 10^6 cells. This error has been verified and does not affect the conclusion of the paper. The correct graph is shown below in Fig. 1. The authors would like to apologize for any confusion the error may have caused.

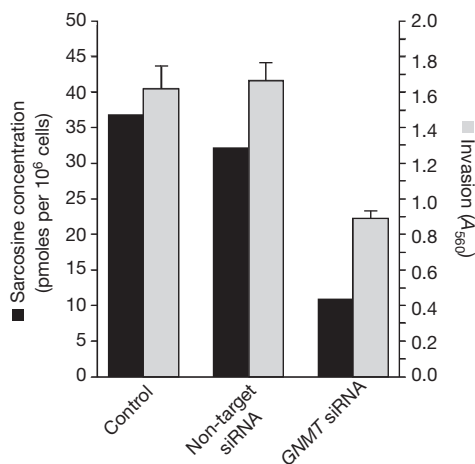


Figure 1 | This is the corrected Fig. 4b of the original Article.

mTORC1 couples immune signals and metabolic programming to establish T_{reg}-cell function

Hu Zeng¹, Kai Yang¹, Caryn Cloer¹, Geoffrey Neale², Peter Vogel³ & Hongbo Chi¹

The mechanistic target of rapamycin (mTOR) pathway integrates diverse environmental inputs, including immune signals and metabolic cues, to direct T-cell fate decisions¹. The activation of mTOR, which is the catalytic subunit of the mTORC1 and mTORC2 complexes, delivers an obligatory signal for the proper activation and differentiation of effector CD4⁺ T cells^{2,3}, whereas in the regulatory T-cell (T_{reg}) compartment, the Akt–mTOR axis is widely acknowledged as a crucial negative regulator of T_{reg}-cell *de novo* differentiation^{4–8} and population expansion⁹. However, whether mTOR signalling affects the homeostasis and function of T_{reg} cells remains largely unexplored. Here we show that mTORC1 signalling is a pivotal positive determinant of T_{reg}-cell function in mice. T_{reg} cells have elevated steady-state mTORC1 activity compared to naive T cells. Signals through the T-cell antigen receptor (TCR) and interleukin-2 (IL-2) provide major inputs for mTORC1 activation, which in turn programs the suppressive function of T_{reg} cells. Disruption of mTORC1 through T_{reg}-specific deletion of the essential component raptor leads to a profound loss of T_{reg}-cell suppressive activity *in vivo* and the development of a fatal early onset inflammatory disorder. Mechanistically, raptor/mTORC1 signalling in T_{reg} cells promotes cholesterol and lipid metabolism, with the mevalonate pathway particularly important for coordinating T_{reg}-cell proliferation and upregulation of the suppressive molecules CTLA4 and ICOS to establish T_{reg}-cell functional competency. By

contrast, mTORC1 does not directly affect the expression of Foxp3 or anti- and pro-inflammatory cytokines in T_{reg} cells, suggesting a non-conventional mechanism for T_{reg}-cell functional regulation. Finally, we provide evidence that mTORC1 maintains T_{reg}-cell function partly through inhibiting the mTORC2 pathway. Our results demonstrate that mTORC1 acts as a fundamental ‘rheostat’ in T_{reg} cells to link immunological signals from TCR and IL-2 to lipogenic pathways and functional fitness, and highlight a central role of metabolic programming of T_{reg}-cell suppressive activity in immune homeostasis and tolerance.

The evolutionarily conserved mTOR signalling pathway couples cell growth and proliferation to nutrient availability and metabolic cues¹⁰. To investigate the function of mTORC1 in naturally occurring T_{reg} cells, we compared mTORC1 activity between T_{reg} cells and naive T cells at steady state. T_{reg} cells had relatively increased phosphorylation of S6 and 4E-BP1, two major mTORC1 downstream targets (Fig. 1a and Supplementary Fig. 1a–c), whereas STAT5 phosphorylation was similar in both sets of cells (Supplementary Fig. 1d). This finding is consistent with a recent study describing elevated S6 phosphorylation in T_{reg} cells compared with non-T_{reg} cells¹¹. T_{reg} cells also contained a higher abundance of CD71 (the transferrin receptor) and, to a lesser extent, CD98 (a subunit of the L-amino acid transporter), key nutrient receptors that depend on mTORC1 activity for expression¹² (Fig. 1b).

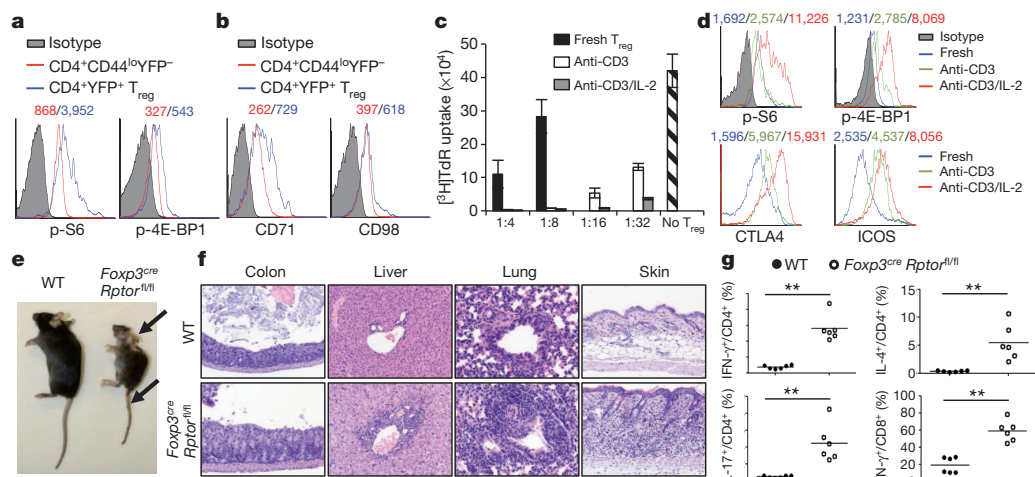


Figure 1 | mTORC1 signalling is constitutively active in T_{reg} cells and its disruption results in a fatal early onset inflammatory disorder. **a**, Comparison of phosphorylation (p) of S6 and 4E-BP1 between naive T cells (CD4⁺CD44^{lo}Foxp3[−]) and T_{reg} cells (CD4⁺Foxp3⁺; from Foxp3^{cre} mice). **b**, Comparison of CD71 and CD98 expression between naive T cells and T_{reg} cells. **a**, **b**, Mean fluorescent intensity (MFI) is presented above the plots. **c**, T_{reg} cells from C57BL/6J mice were activated with anti-CD3 or anti-CD3 and IL-2 for 3 days, washed and then used for *in vitro* suppression assays at multiple T_{reg} versus T_{eff} ratios; freshly isolated T_{reg} cells were included for comparison (shown only for 1:4 and 1:8 ratios). Error bars represent standard deviation

(s.d.) ($n = 3$). **d**, Comparison of phosphorylation of S6 and 4E-BP1 and expression of CTLA4 and ICOS between freshly isolated and pre-activated T_{reg} cells. MFI is presented above the plots. **e**, Images of 42-day-old wild-type (WT) and Foxp3^{cre}Raptor^{fl/fl} mice. Arrows indicate the scaly tail and ulceration of the body. **f**, Haematoxylin and eosin staining of colon (original magnification, $\times 20$), lung ($\times 40$), liver ($\times 20$) and skin ($\times 20$) from 21-day-old wild-type and Foxp3^{cre}Raptor^{fl/fl} mice. **g**, IFN- γ , IL-17 and IL-4 production in CD4⁺ cells and IFN- γ production in CD8⁺ cells from wild-type and Foxp3^{cre}Raptor^{fl/fl} mice. P values are determined by Mann–Whitney test. ** $P < 0.01$. Results represent five (**a**, **g**), three (**b**) and two (**c**–**f**) independent experiments.

¹Department of Immunology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ²Hartwell Center for Bioinformatics and Biotechnology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ³Department of Pathology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA.

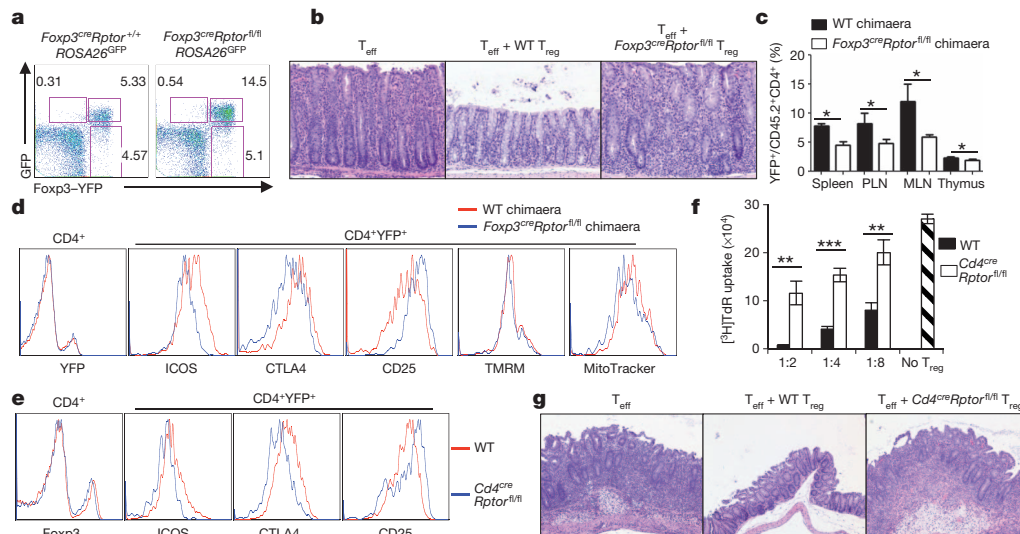


Figure 2 | Raptor deletion abrogates T_{reg} -cell suppressive activity. **a**, Foxp3-YFP and green fluorescent protein (GFP) expression in $CD4^{+}$ T cells from 25-day-old $Foxp3^{cre}Raptor^{+/+}/ROSA26^{GFP}$ and $Foxp3^{cre}Raptor^{fl/fl}/ROSA26^{GFP}$ mice. **b**, Representative colon histology in $Rag1^{-/-}$ mice given T_{eff} cells alone or in combination with wild-type (WT) or $Foxp3^{cre}Raptor^{fl/fl}$ T_{reg} cells (original magnification, $\times 20$). **c**, Percentages of T_{reg} cells in mixed chimaeras reconstituted with bone marrow cells from $CD45.1^{+}$ mice and wild-type or $Foxp3^{cre}Raptor^{fl/fl}$ mice ($n = 3$). MLN, mesenteric lymph nodes; PLN, peripheral lymph nodes. **d**, Expression of T_{reg} signature molecules and mitochondrial membrane potential (tetramethylrhodamine methyl ester (TMRM)) and mass

(MitoTracker) in the reconstituted chimaeras. **e**, Expression of T_{reg} signature molecules in 6-week-old wild-type and $Cd4^{cre}Raptor^{fl/fl}$ mice. **f**, *In vitro* suppression assays mediated by T_{reg} cells from wild-type and $Cd4^{cre}Raptor^{fl/fl}$ mice. Error bars represent s.d. ($n = 3$). **g**, Representative colon histology in $Rag1^{-/-}$ mice given T_{eff} cells alone or in combination with wild-type or $Cd4^{cre}Raptor^{fl/fl}$ T_{reg} cells (original magnification, $\times 20$). *P* values are determined by Mann-Whitney test (**c**) or *t*-test (**f**). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. Results represent three (**a**, **d**, **f**), two (**b**, **c**, **g**) and five (**e**) independent experiments.

Previous studies have demonstrated a requirement of mTORC1 for mitochondrial metabolism¹³, the dysregulation of which could affect the homeostasis of memory T cells¹⁴ and haematopoietic stem cells¹⁵. T_{reg} cells had reduced mitochondrial membrane potential, whereas their mitochondrial mass and reactive oxygen species (ROS) production were largely comparable to naive cells (Supplementary Fig. 1e). Thus, T_{reg} cells exhibit distinct regulation of mTORC1 activity and metabolism under steady-state conditions.

One of the hallmarks of T_{reg} cells is that they are anergic *in vitro* on TCR stimulation alone¹⁶, but are highly proliferative *in vivo*^{17,18}. Whereas reversal of anergy has been suggested to impair T_{reg} function¹⁶, T_{reg} cells stimulated by TCR together with IL-2, or under lymphopenic conditions, possess enhanced suppressive activity^{18,19}. Indeed, prior stimulation of T_{reg} cells with anti-CD3 upregulated their suppressive activity, and this was further enhanced by concomitant treatment with IL-2 (Fig. 1c). Such functional enhancements were associated with increased mTORC1 activity and expression of CTLA4 and ICOS, important effector molecules mediating T_{reg} function^{20,21} (Fig. 1d). Compared with IL-2, CD28-mediated co-stimulation had more modest enhancing effects on these events (Supplementary Fig. 1f, g). Thus, TCR and IL-2 are the predominant signals that promote mTORC1 activity and the suppressive function of T_{reg} cells.

Although these results indicate that mTORC1 is positively correlated with T_{reg} suppressive activity, a number of recent studies have revealed a negative role of mTORC1 in T_{reg} cells, including in the suppression of T_{reg} differentiation^{2,4–8} and population expansion⁹. To establish the physiological relevance of mTORC1 signalling in T_{reg} cells, we deleted raptor, an obligatory component of the mTORC1 complex¹⁰, in T_{reg} cells by crossing $Raptor^{fl/fl}$ mice with $Foxp3^{YFP-Cre}$ mice²² (designated $Foxp3^{cre}Raptor^{fl/fl}$). Raptor expression and S6 and 4E-BP1 phosphorylation were abolished in raptor-deficient T_{reg} cells, but were normal in $Foxp3^{-}$ populations (Supplementary Fig. 2a–c). Additionally, a slight increase of phosphorylation of Akt Ser473 was observed in mutant T_{reg} cells, whereas phosphorylation of Akt Thr308 and Erk in T_{reg} cells was not affected (Supplementary Fig. 2d). Thus, deletion of raptor specifically abrogated mTORC1 activity.

Surprisingly, $Foxp3^{cre}Raptor^{fl/fl}$ mice developed profound inflammatory diseases, indicated by a reduced body size and hunched posture, crusting of ears, eyelids and tail, and skin ulceration particularly on the head and upper back (Fig. 1e). $Foxp3^{cre}Raptor^{fl/fl}$ mice showed extensive lymphadenopathy (Supplementary Fig. 3a) and infiltrations of lymphocytes and myeloid cells in colon mucosa, lung, liver sinusoids and skin (Fig. 1f), among other organs, before they died at a young age (Supplementary Fig. 3b). $Foxp3^{cre}Raptor^{fl/fl}$ mice had an increased $CD8^{+}$ -cell percentage and altered $CD4^{+}/CD8^{+}$ ratio, highly elevated memory/effector phenotype ($CD44^{hi}CD62L^{lo}$) T cells, and expansion of $CD11b^{+}$ granulocytes and macrophages (Supplementary Fig. 3c–e). Furthermore, T cells from $Foxp3^{cre}Raptor^{fl/fl}$ mice showed increased interferon (IFN)- γ -, IL-4- and IL-17-producing $CD4^{+}$ cells and IFN- γ -producing $CD8^{+}$ cells (Fig. 1g). These phenotypes are reminiscent of those observed in mice carrying the scurfy mutation²³, indicating a loss of T_{reg} function.

Despite developing severe autoimmune diseases, $Foxp3^{cre}Raptor^{fl/fl}$ mice had increased numbers of $Foxp3^{+}$ T_{reg} cells (Supplementary Fig. 4a). Lineage-tracing experiments indicated that the increase of T_{reg} cells was mainly ascribed to a more abundant population of canonical T_{reg} cells, as indicated by concomitant expression of yellow fluorescent protein (YFP)-tagged Foxp3 and specific Cre activity (Fig. 2a). Raptor-deficient T_{reg} cells had normal Foxp3 expression and either increased or normal expression of many T_{reg} signature molecules examined, whereas CD62L expression was reduced (Supplementary Fig. 4b). Thus, in settings of immune activation, raptor-deficient T_{reg} cells acquired activated phenotypes. Additional analysis revealed highly reduced mitochondrial membrane potential and mass, and ROS production, in mutant T_{reg} cells, whereas *in vitro* suppressive activity seemed to be largely undisturbed (Supplementary Fig. 4b, c). To test the function of raptor-deficient T_{reg} cells *in vivo*, we used a model of colitis induced by the transfer of pathogenic effector T cells (T_{eff} ; $CD4^{+}CD45RB^{hi}CD25^{-}$) into $Rag1^{-/-}$ hosts, which could be prevented by cotransfer with wild-type T_{reg} cells. However, cotransfer with raptor-deficient T_{reg} cells failed to inhibit colitis (Fig. 2b), or the expansion and IFN- γ production of T_{eff} cells (Supplementary Fig. 4d). Therefore, raptor is required for the suppressive function of T_{reg} cells *in vivo*.

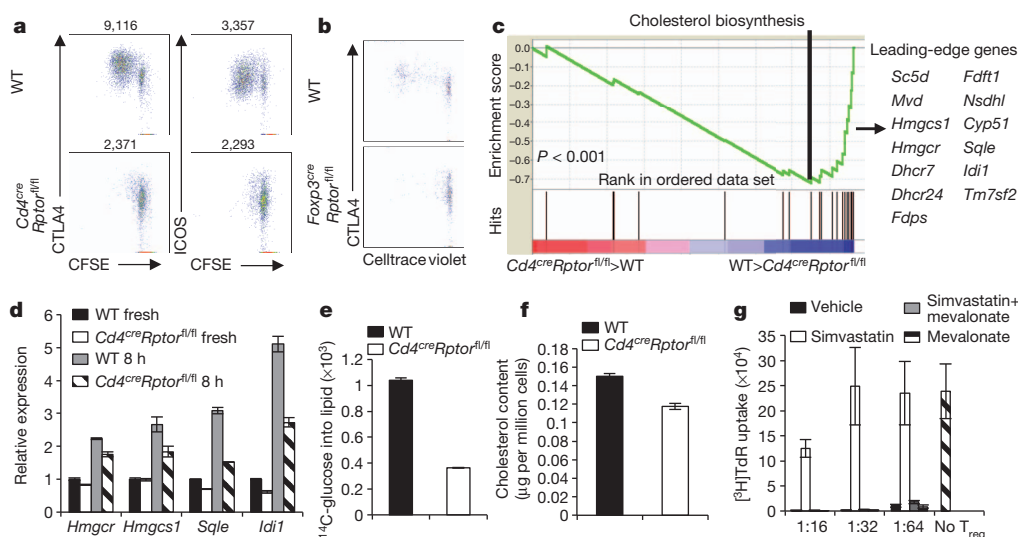


Figure 3 | Raptor coordinates T_{reg}-cell proliferation and effector molecule expression by orchestrating cholesterol/lipid biosynthetic metabolism, especially the mevalonate pathway. **a**, Carboxyfluorescein diacetate succinimidyl ester (CFSE)-labelled T_{reg} cells from wild-type (WT) and *Cd4^{cre}Raptor^{fl/fl}* mice were stimulated with anti-CD3, anti-CD28 and IL-2 for 3 days, followed by analysis for CFSE dilution and CTLA4 and ICOS expression (MFI presented above the plots). **b**, Celltrace violet-labelled T_{reg} cells from wild-type and *Foxp3^{cre}Raptor^{fl/fl}* mice were transferred into unmanipulated CD45.1⁺ mice, and Celltrace violet dilution and CTLA4 expression were analysed 7 days later. **c**, Gene-set enrichment analysis revealed the underrepresentation of cholesterol biosynthesis genes in raptor-deficient T_{reg} cells. The lower part of the plot shows the distribution of the genes in the cholesterol biosynthesis

signature gene set ('Hits') against the ranked list of genes, and the list on the right shows the top hit genes. **d**, Real-time PCR quantification of lipogenic gene expression in wild-type and *Cd4^{cre}Raptor^{fl/fl}* freshly isolated T_{reg} cells or those stimulated with anti-CD3, anti-CD28 and IL-2 for 8 h. **e**, *De novo* lipogenesis of T_{reg} cells from wild-type and *Cd4^{cre}Raptor^{fl/fl}* mice after 24 h of anti-CD3, anti-CD28 and IL-2 stimulation, with the final 4 h labelled with ¹⁴C-glucose. **f**, Total cholesterol level was measured using lysates of activated T_{reg} cells from wild-type and *Cd4^{cre}Raptor^{fl/fl}* mice. **g**, *In vitro* suppression assays mediated by T_{reg} cells previously activated by anti-CD3, anti-CD28 and IL-2 for 3 days in the presence of vehicle, simvastatin (2 μM), simvastatin and mevalonate (100 μM) or mevalonate alone. Error bars represent s.d. (*n* = 3). Results represent two independent experiments except for **c** (*n* = 4 from one experiment).

We speculated that the increased number and heightened activation phenotypes of T_{reg} cells in *Foxp3^{cre}Raptor^{fl/fl}* mice could be due to a compensatory response to the continuing inflammation²⁴. To test this, we generated mixed bone marrow chimaeras by reconstituting sublethally irradiated *Rag1^{-/-}* mice with bone marrow cells from CD45.1⁺ mice mixed with those from either wild-type or *Foxp3^{cre}Raptor^{fl/fl}* mice. Raptor-deficient T_{reg} cells were underrepresented in the chimaeras (Fig. 2c), with impaired expression of CTLA4 and ICOS but largely normal CD25 levels and mitochondrial parameters (Fig. 2d). These findings were recapitulated in 10-day-old *Foxp3^{cre}Raptor^{fl/fl}* mice, which exhibited minimal immune activation (Supplementary Fig. 4e). Thus, in a disease-free environment, loss of raptor in T_{reg} cells renders them at a competitive disadvantage and compromises the expression of CTLA4 and ICOS. By contrast, the pronounced mitochondrial defects in T_{reg} cells from *Foxp3^{cre}Raptor^{fl/fl}* mice (Supplementary Fig. 4b) were not intrinsically due to raptor deficiency, but were probably a compensatory response, as observed in haematopoietic stem cells lacking the metabolic sensor LKB1 (ref. 15).

To facilitate mechanistic studies of raptor-deficient T_{reg} cells without the influence of continuing inflammation, we generated *Cd4-Cre;Raptor^{fl/fl}* (*Cd4^{cre}Raptor^{fl/fl}*) mice, in which raptor was deleted in all αβ T cells. *Cd4^{cre}Raptor^{fl/fl}* mice showed a small reduction of T_{reg} numbers in peripheral lymph nodes but not the spleen (Supplementary Fig. 5a), and immunoblots confirmed the loss of S6 and 4E-BP1 phosphorylation in these T_{reg} cells (Supplementary Fig. 5b). Expression of CTLA4, ICOS and CD71 was impaired in T_{reg} cells from *Cd4^{cre}Raptor^{fl/fl}* mice, whereas the levels of Foxp3, CD25 and other markers were normal (Fig. 2e and Supplementary Fig. 5c). *In vitro* assays revealed a loss of the suppressive activity of T_{reg} cells from *Cd4^{cre}Raptor^{fl/fl}* mice (Fig. 2f). These T_{reg} cells also exhibited severely impaired *in vivo* suppressive activity, because they failed to suppress colitis or IFN-γ production mediated by T_{eff} cells (Fig. 2g and Supplementary Fig. 5d). Analysis of mixed chimaeras composed of CD45.1⁺ and *Cd4^{cre}Raptor^{fl/fl}* cells confirmed that raptor-deficient T_{reg} cells were reduced as a percentage of total CD4⁺ T cells with

impaired CTLA4 and ICOS expression (Supplementary Fig. 5e). Thus, T_{reg} cells deficient in mTORC1 have an intrinsic defect in suppressive activity. Nonetheless, *Cd4^{cre}Raptor^{fl/fl}* mice exhibited normal health status, which was probably due to impaired activation of conventional T cells (K.Y. and H.C., unpublished observations). We found a similar requirement of raptor in T_{reg} suppressive activity after Cre-ER^{T2}-mediated acute deletion of raptor in adult mice (Supplementary Fig. 6).

Cytokines produced by T_{reg} cells represent an important immunosuppressive mechanism²⁵. The anti-inflammatory cytokines TGF-β1 and IL-10 were expressed normally in raptor-deficient T_{reg} cells (Supplementary Fig. 7a, b). Although T_{reg} cells from *Foxp3^{cre}Raptor^{fl/fl}* mice expressed increased IFN-γ, which may compromise T_{reg} function²⁶, T_{reg} cells from *Cd4^{cre}Raptor^{fl/fl}* did not (data not shown), thereby excluding such phenotypes as cell intrinsic. Moreover, ablation of IFN-γ did not ameliorate the immune-activation phenotypes of *Foxp3^{cre}Raptor^{fl/fl}* mice (Supplementary Fig. 7c). Thus, the severe autoimmune phenotypes of *Foxp3^{cre}Raptor^{fl/fl}* mice probably occur independently of cytokine dysregulation.

We therefore focused on the effects of raptor on T_{reg} surface effector molecules, especially CTLA4 and ICOS, which were greatly reduced in T_{reg} cells from multiple raptor-deficient genetic models in an intrinsic manner (Fig. 2d, e and Supplementary Figs 4e, 5e). Moreover, CTLA4 and, to a lesser extent, ICOS were further increased upon *in vitro* stimulation, and this upregulation correlated with T_{reg}-cell proliferation. Such upregulation was blunted in T_{reg} cells from *Cd4^{cre}Raptor^{fl/fl}* mice, which was associated with impaired proliferation (Fig. 3a and Supplementary Fig. 8a). Moreover, when wild-type T_{reg} cells were transferred into unmanipulated CD45.1⁺ mice, a proportion of them divided and upregulated CTLA4 (Fig. 3b). However, T_{reg} cells from *Cd4^{cre}Raptor^{fl/fl}* or *Foxp3^{cre}Raptor^{fl/fl}* mice failed to proliferate or upregulate CTLA4 on transfer into CD45.1⁺ mice (Fig. 3b and Supplementary Fig. 8b). Markedly diminished homeostatic proliferation and CTLA4 upregulation were also observed when raptor-deficient T_{reg} cells were transferred into *Rag1^{-/-}* mice (Supplementary Fig. 8c). Altogether, loss of raptor inhibits T_{reg}-cell proliferation *in vitro* and *in vivo*. Furthermore,

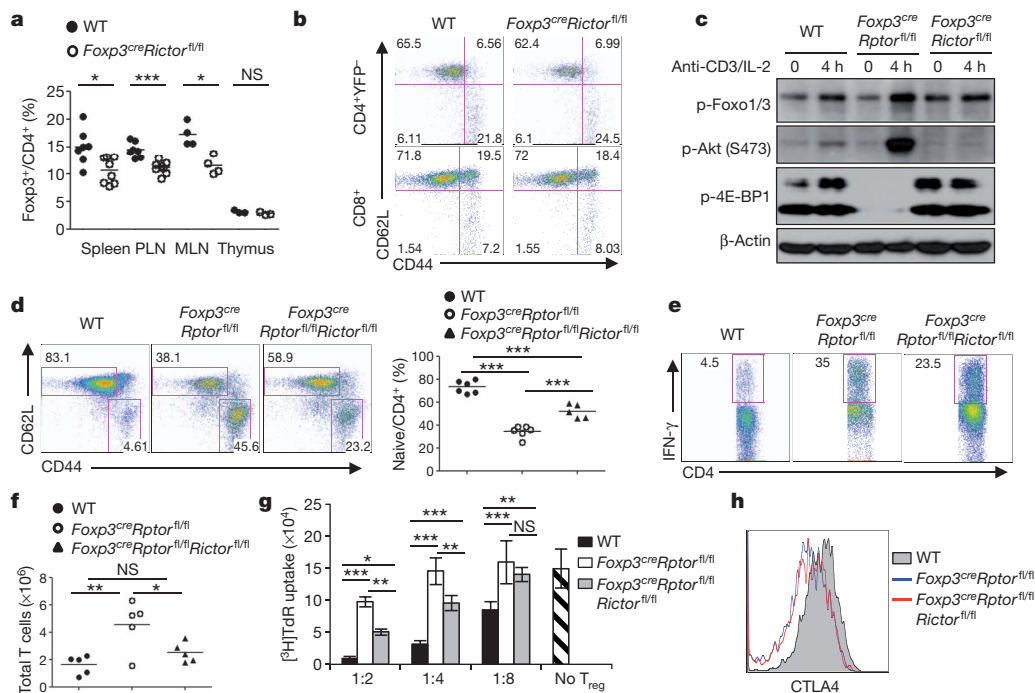


Figure 4 | Deletion of *Rictor* does not alter T_{reg}-cell function but partially rescues inflammation in *Foxp3^{cre}Rictor^{fl/fl}* mice. **a**, T_{reg} percentage in 2–3-month-old wild-type (WT) and *Foxp3^{cre}Rictor^{fl/fl}* mice. NS, not significant. **b**, Expression of CD62L and CD44 on splenic T cells from wild-type and *Foxp3^{cre}Rictor^{fl/fl}* mice. **c**, T_{reg} cells were activated by anti-CD3 and IL-2 for 4 h followed by immunoblots. **d**, Expression of CD62L and CD44 on splenic CD4⁺ T cells from 3–4-week-old wild-type, *Foxp3^{cre}Rictor^{fl/fl}* and *Foxp3^{cre}Rictor^{fl/fl}* mice. Right, percentage of CD62L^{hi}CD44^{lo} naive CD4⁺ cells in the spleen. **e**, IFN-γ production in splenic CD4⁺ T cells from

wild-type, *Foxp3^{cre}Rictor^{fl/fl}* and *Foxp3^{cre}Rictor^{fl/fl}* mice. **f**, Numbers of total TCRβ⁺ cells in peripheral lymph nodes from wild-type, *Foxp3^{cre}Rictor^{fl/fl}* and *Foxp3^{cre}Rictor^{fl/fl}* mice. **g**, *In vitro* suppression assays mediated by T_{reg} cells from 3–4-week-old wild-type, *Cd4^{cre}Rictor^{fl/fl}* and *Cd4^{cre}Rictor^{fl/fl}* mice. Error bars represent s.d. (*n* = 3). **h**, CTLA4 expression in T_{reg} cells from the spleen of wild-type, *Cd4^{cre}Rictor^{fl/fl}* and *Cd4^{cre}Rictor^{fl/fl}* mice. *P* values are determined by Mann–Whitney test (**a**) and analysis of variance (**d**, **f**, **g**). **P* < 0.05; ***P* < 0.01; ****P* < 0.001. Results represent three (**a**–**f**) and two (**g**, **h**) independent experiments.

the proliferative defect was associated specifically with impaired CTLA4 and ICOS expression, whereas apoptosis or other activation-induced events such as upregulation of GATA3 and CD103 were independent of raptor (data not shown). Thus, mTORC1 is important for coordinating T_{reg}-cell proliferation and CTLA4 and ICOS expression, which may collectively contribute to the T_{reg} suppressive function.

To explore raptor/mTORC1-dependent molecular mechanisms, we used functional genomics and compared gene expression profiles of T_{reg} cells from *Cd4^{cre}Rictor^{fl/fl}* and wild-type mice. Gene-set enrichment analysis revealed that cholesterol biosynthesis was the most altered pathway in a raptor-dependent manner: expression of many genes in this pathway was downregulated after raptor deletion (Fig. 3c and Supplementary Fig. 9a), including those encoding 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR), squalene epoxidase (SQLE) and isopentenyl-diphosphate delta isomerase 1 (IDI1), which was verified by real-time polymerase chain reaction (PCR) analysis (Fig. 3d). Furthermore, on TCR stimulation, raptor-deficient T_{reg} cells failed to synthesize lipids from glucose efficiently (Fig. 3e), and had reduced cellular cholesterol (Fig. 3f). These results establish a role of mTORC1 in promoting cholesterol/lipid biosynthesis. Given the interconnected relationship between the biosynthetic and bioenergetic programs²⁷, we determined the role of mTORC1 in T_{reg} bioenergetics by measuring the extracellular acidification rate (ECAR) and oxygen consumption rate (OCR), which denoted glycolysis and mitochondrial respiration, respectively. Raptor-deficient T_{reg} cells showed decreased ECAR and OCR (Supplementary Fig. 9b, c), indicating a requirement for mTORC1 in bioenergetics. We therefore determined the *de novo* lipid synthesis rate by measuring the incorporation of ¹⁴C-acetate into cellular lipids, thereby bypassing the requirement of mTORC1 in glycolysis or mitochondrial activity. Indeed, raptor deficiency diminished the synthesis of lipids from ¹⁴C-acetate in T_{reg} cells (Supplementary

Fig. 9d). These results demonstrated a specific role of mTORC1 in orchestrating the lipogenic program.

To test the functional significance of lipid metabolism, we activated T_{reg} cells in the presence of 25-hydroxycholesterol, and this general lipid synthesis inhibitor potently blocked T_{reg} suppressive activity (Supplementary Fig. 10a). Direct inhibition of HMGCR, the rate-limiting enzyme in the synthesis of cholesterol and isoprenoid lipids, by simvastatin also impaired T_{reg} suppressive activity. Importantly, simvastatin-induced inhibition was completely reversed by the simultaneous addition of mevalonate, the metabolite downstream of HMGCR (Fig. 3g). Similar effects were observed after treatments with atorvastatin and lovastatin (Supplementary Fig. 10b). The inhibitory effects of these agents on T_{reg}-cell function were associated with impaired T_{reg}-cell proliferation and effector molecule upregulation (Supplementary Fig. 10c) in a mevalonate-dependent manner (Supplementary Fig. 10d). Furthermore, proliferation and CTLA4 and ICOS upregulation in T_{reg} cells transferred into congenic CD45.1⁺ mice were diminished by statin treatment *in vivo* (Supplementary Fig. 11). Altogether, our results show that raptor/mTORC1 signalling promotes the lipogenic program, with the mevalonate pathway particularly important for mediating T_{reg}-cell proliferation, CTLA4 and ICOS upregulation, and functional fitness.

Moreover, lipid-metabolism-dependent regulation of CTLA4 and ICOS is also operative in conventional T cells, because naive T cells treated with inhibitors of this pathway or deficient in raptor failed to upregulate these molecules effectively after TCR stimulation (Supplementary Fig. 12). We speculate that this constitutes a feedback mechanism, which would account for the immunosuppressive activity of activated non-T_{reg} cells previously described^{28,29}, a notion that awaits further investigation.

Crosstalk between the mTORC1 and mTORC2 signalling pathways has been reported¹⁰, although the molecular details and functional

significance are unclear. Notably, mTORC2 contributes to Akt activation, which in turn signals to Foxo1 (ref. 26), and acts in synergy with mTORC1 to negatively control the generation of induced T_{reg} cells². We therefore examined the role of mTORC2 and its interaction with mTORC1 in naturally occurring T_{reg} cells. T_{reg} -specific deletion of *Rictor*, the defining component of mTORC2, resulted in a small reduction of T_{reg} cells in the periphery (Fig. 4a). However, loss of Rictor in T_{reg} cells did not perturb T-cell homeostasis, cellularity or cytokine expression at steady state (Fig. 4b and Supplementary Fig. 13a, b), or the health status of the mice. Also, Rictor-deficient T_{reg} cells had normal expression of T_{reg} signature molecules, mitochondrial activity and *in vitro* suppressive activity (Supplementary Fig. 13c, d). Thus, in contrast to a crucial requirement of mTORC1 in programming T_{reg} activity, mTORC2 probably contributes to T_{reg} maintenance but is dispensable for T_{reg} function.

To examine the crosstalk between the two mTOR complexes, we analysed mTORC2 activity in raptor-deficient T_{reg} cells. Phosphorylation of well-established mTORC2 targets, Akt Ser473 and Foxo1/3, was elevated in raptor-deficient T_{reg} cells after anti-CD3/IL-2 (Fig. 4c) or anti-CD3/CD28 stimulation (Supplementary Fig. 14), consistent with the observation in freshly isolated raptor-deficient T_{reg} cells (Supplementary Fig. 2d). By contrast, Akt Ser473 phosphorylation was nearly abolished after *Rictor* deletion (Fig. 4c and Supplementary Fig. 14a). Thus, mTORC1 negatively feeds back on mTORC2 in T_{reg} cells. To determine the contribution of mTORC2 activity to *Foxp3^{cre}Rptor^{fl/fl}* phenotypes, we generated *Foxp3^{cre}Rptor^{fl/fl}Rictor^{fl/fl}* mice. Compared with T cells in *Foxp3^{cre}Rptor^{fl/fl}* mice, those in *Foxp3^{cre}Rptor^{fl/fl}Rictor^{fl/fl}* mice contained an increased percentage of naive phenotypes (Fig. 4d) and showed modestly reduced IFN- γ production (Fig. 4e). Also, the lymphadenopathy phenotype observed in *Foxp3^{cre}Rptor^{fl/fl}* mice was ameliorated in *Foxp3^{cre}Rptor^{fl/fl}Rictor^{fl/fl}* mice (Fig. 4f). To test the intrinsic suppressive activity of raptor- and Rictor-deficient T_{reg} cells, we generated *Cd4^{cre}Rptor^{fl/fl}Rictor^{fl/fl}* mice that eliminated both mTORC1 and mTORC2 activities (Supplementary Fig. 15a, b). An *in vitro* suppression assay revealed an intermediate phenotype of raptor/Rictor-deficient T_{reg} cells compared with raptor-deficient and wild-type cells (Fig. 4g). However, impaired CTLA4 expression and TCR-induced proliferation after raptor deletion were not restored by the concomitant loss of Rictor (Fig. 4h and Supplementary Fig. 15c). Furthermore, *Foxp3^{cre}Rptor^{fl/fl}Rictor^{fl/fl}* mice still succumbed to an inflammatory disorder, albeit with a small extension of lifespan and less inflammation as compared with *Foxp3^{cre}Rptor^{fl/fl}* mice (Supplementary Fig. 16). Therefore, Raptor/mTORC1 signalling promotes T_{reg} function in part through inhibiting mTORC2, although the predominant effect of mTORC1 is to coordinate metabolic programs and T_{reg} suppressive molecules.

Much emphasis has been placed on the transcriptional mechanisms that orchestrate T_{reg} suppressive activity, but how immunological signals are sensed and integrated by T_{reg} cells for their functional activation remains obscure. We show that mTORC1-dependent metabolic programming is a central mechanism to couple immune signals, including TCR and IL-2, and T_{reg} suppressive function. Raptor/mTORC1 promotes the lipogenic program, with the mevalonate pathway particularly important for coordinating T_{reg} proliferation and the optimal induction of effector molecules CTLA4 and ICOS (Supplementary Fig. 17). Thus, mTORC1-dependent lipid metabolism provides a novel link between two crucial T_{reg} -cell regulators, IL-2 and CTLA4, the loss of which causes fulminant autoimmune diseases^{21,30}. An additional mechanism is through the mTORC1-dependent inhibition of mTORC2, although the modest rescue effect of *Rictor* deletion in *Foxp3^{cre}Rptor^{fl/fl}* mice suggests a relatively minor contribution of this feedback pathway. Notably, mTORC1-dependent metabolic programming operates both under steady state and immune stimulation, and this provides important mechanistic insights into the apparently conflicting observations of the *in vitro* anergy of T_{reg} cells and their *in vivo* antigen-primed state^{16–18}. We propose that T_{reg} cells adopt the evolutionarily ancient mTORC1

signalling pathway to link immunological inputs to metabolic activity and functional fitness, therefore implicating mTORC1 as a fundamental rheostat to program T_{reg} -cell suppressive activity through a non-conventional mechanism.

METHODS SUMMARY

The mice used in this study were backcrossed onto the C57BL/6J background. Bone marrow chimeras were generated by transferring 1×10^7 T-cell-depleted bone marrow cells into sub-lethally irradiated (5 Gy) *Rag1^{-/-}* mice, followed by reconstitution for at least 2 months. T-cell-mediated colitis was induced by intraperitoneal transfer of 4×10^5 T_{eff} cells ($CD4^+CD45RB^{hi}CD25^-$) from $CD45.1^+$ mice in the absence or presence of 2×10^5 T_{reg} cells into *Rag1^{-/-}* mice. Lymphocytes were stained with surface antibodies or intracellularly for Foxp3, CTLA4 and caspase 3. Staining for the phosphorylated signalling proteins was carried out with Phosflow kits (BD Biosciences). T_{reg} cells were activated with plate-coated anti-CD3, anti-CD28 and IL-2, in the presence of simvastatin (2 μ M; EMD Millipore), 25-hydroxycholesterol (0.5 μ g ml⁻¹; Sigma), or vehicle. The *de novo* lipogenesis assay was performed by incubating activated T_{reg} cells with ¹⁴C-glucose or ¹⁴C-acetate for 4 h, followed by cell lysis and lipid extraction.

Full Methods and any associated references are available in the online version of the paper.

Received 23 December 2012; accepted 13 May 2013.

Published online 30 June 2013.

- Chi, H. Regulation and function of mTOR signalling in T cell fate decisions. *Nature Rev. Immunol.* **12**, 325–338 (2012).
- Delgoffe, G. M. *et al.* The kinase mTOR regulates the differentiation of helper T cells through the selective activation of signaling by mTORC1 and mTORC2. *Nature Immunol.* **12**, 295–303 (2011).
- Lee, K. *et al.* Mammalian target of rapamycin protein complex 2 regulates differentiation of Th1 and Th2 cell subsets via distinct signaling pathways. *Immunity* **32**, 743–753 (2010).
- Delgoffe, G. M. *et al.* The mTOR kinase differentially regulates effector and regulatory T cell lineage commitment. *Immunity* **30**, 832–844 (2009).
- Haxhinasto, S., Mathis, D. & Benoist, C. The AKT–mTOR axis regulates *de novo* differentiation of $CD4^+$ Foxp3⁺ cells. *J. Exp. Med.* **205**, 565–574 (2008).
- Sauer, S. *et al.* T cell receptor signaling controls Foxp3 expression via PI3K, Akt, and mTOR. *Proc. Natl Acad. Sci. USA* **105**, 7797–7802 (2008).
- Liu, G. *et al.* The receptor S1P₁ overrides regulatory T cell-mediated immune suppression through Akt–mTOR. *Nature Immunol.* **10**, 769–777 (2009).
- Liu, G., Yang, K., Burns, S., Shrestha, S. & Chi, H. The S1P₁–mTOR axis directs the reciprocal differentiation of T_H1 and T_{reg} cells. *Nature Immunol.* **11**, 1047–1056 (2010).
- Battaglia, M., Stabilini, A. & Roncarolo, M. G. Rapamycin selectively expands $CD4^+CD25^+$ Foxp3⁺ regulatory T cells. *Blood* **105**, 4743–4748 (2005).
- Laplanche, M. & Sabatini, D. M. mTOR signaling in growth control and disease. *Cell* **149**, 274–293 (2012).
- Procaccini, C. *et al.* An oscillatory switch in mTOR kinase activity sets regulatory T cell responsiveness. *Immunity* **33**, 929–941 (2010).
- Kelly, A. P. *et al.* Notch-induced T cell development requires phosphoinositide-dependent kinase 1. *EMBO J.* **26**, 3441–3450 (2007).
- Cunningham, J. T. *et al.* mTOR controls mitochondrial oxidative function through a YY1–PGC-1 α transcriptional complex. *Nature* **450**, 736–740 (2007).
- van der Windt, G. J. *et al.* Mitochondrial respiratory capacity is a critical regulator of $CD8^+$ T cell memory development. *Immunity* **36**, 68–78 (2012).
- Gurumurthy, S. *et al.* The Lkb1 metabolic sensor maintains haematopoietic stem cell survival. *Nature* **468**, 659–663 (2010).
- Takahashi, T. *et al.* Immunologic self-tolerance maintained by $CD25^+$ $CD4^+$ naturally anergic and suppressive T cells: induction of autoimmune disease by breaking their anergic/suppressive state. *Int. Immunol.* **10**, 1969–1980 (1998).
- Fisson, S. *et al.* Continuous activation of autoreactive $CD4^+$ $CD25^+$ regulatory T cells in the steady state. *J. Exp. Med.* **198**, 737–746 (2003).
- Gavin, M. A., Clarke, S. R., Negrou, E., Gallegos, A. & Rudensky, A. Homeostasis and anergy of $CD4^+$ $CD25^+$ suppressor T cells *in vivo*. *Nature Immunol.* **3**, 33–41 (2002).
- Thornton, A. M., Donovan, E. E., Piccirillo, C. A. & Shevach, E. M. Cutting edge: IL-2 is critically required for the *in vitro* activation of $CD4^+$ $CD25^+$ T cell suppressor function. *J. Immunol.* **172**, 6519–6523 (2004).
- Herman, A. E., Freeman, G. J., Mathis, D. & Benoist, C. $CD4^+$ $CD25^+$ T regulatory cells dependent on ICOS promote regulation of effector cells in the prediabetic lesion. *J. Exp. Med.* **199**, 1479–1489 (2004).
- Wing, K. *et al.* CTLA-4 control over Foxp3⁺ regulatory T cell function. *Science* **322**, 271–275 (2008).
- Rubtsov, Y. P. *et al.* Regulatory T cell-derived interleukin-10 limits inflammation at environmental interfaces. *Immunity* **28**, 546–558 (2008).
- Kanangat, S. *et al.* Disease in the scurfy (sf) mouse is associated with overexpression of cytokine genes. *Eur. J. Immunol.* **26**, 161–165 (1996).
- Chaudhry, A. *et al.* Interleukin-10 signaling in regulatory T cells is required for suppression of Th17 cell-mediated inflammation. *Immunity* **34**, 566–578 (2011).

25. Josefowicz, S. Z., Lu, L. F. & Rudensky, A. Y. Regulatory T cells: mechanisms of differentiation and function. *Annu. Rev. Immunol.* **30**, 531–564 (2012).
26. Ouyang, W. *et al.* Novel Foxo1-dependent transcriptional programs control T_{reg} cell function. *Nature* **491**, 554–559 (2012).
27. Wang, R. & Green, D. R. Metabolic checkpoints in activated T cells. *Nature Immunol.* **13**, 907–915 (2012).
28. Tai, X. *et al.* Basis of CTLA-4 function in regulatory and conventional CD4⁺ T cells. *Blood* **119**, 5155–5163 (2012).
29. Corse, E. & Allison, J. P. Cutting edge: CTLA-4 on effector T cells inhibits in trans. *J. Immunol.* **189**, 1123–1127 (2012).
30. Sadlack, B. *et al.* Ulcerative colitis-like disease in mice with a disrupted interleukin-2 gene. *Cell* **75**, 253–261 (1993).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Rudensky for *Foxp3*^{YFP-Cre} mice, D. Green for help with metabolic assays, N. Brydon for animal colony management, Y. Wang for editing of the

manuscript, and the St Jude Immunology FACS core facility for cell sorting. This work was supported by the US National Institutes of Health (K01 AR053573, R21 AI094089, R01 AI101407 and R01 NS064599), the Lupus Research Institute, and the American Lebanese Syrian Associated Charities (all to H.C.).

Author Contributions H.Z. designed and performed experiments, and wrote the manuscript; K.Y. contributed to cellular experiments; C.C. contributed to survival curves and technical support; G.N. performed bioinformatic analyses; P.V. performed histological analysis; H.C. designed experiments, contributed to writing the manuscript, and provided overall direction.

Author Information The microarray data have been deposited in the Gene Expression Omnibus under accession GSE46693. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.C. (hongbo.chi@stjude.org).

METHODS

Mice. C57BL/6J, CD45.1⁺, *Raptor*^{fl}, *Rag1*^{-/-}, *Ifng*^{-/-} and ROSA26^{GFP} (a loxP-site-flanked STOP cassette followed by the GFP-encoding sequence was inserted into the ROSA26 locus) mice were purchased from the Jackson Laboratory. *Rictor*^{fl/fl}, *Cd4*-Cre and ROSA26-Cre-ER^{T2} mice have been described previously³¹. *Foxp3*^{YFP-Cre} mice were a gift from A. Rudensky²². *Foxp3*^{cre}*Raptor*^{fl/fl} mice were used at 3–4-weeks old unless otherwise noted, with the age and gender-matched wild-type mice containing the *Foxp3*^{cre} allele as controls. Other mice were used at 8–10-weeks old unless otherwise noted. Bone marrow chimaeras were generated by transferring 1×10^7 T-cell-depleted bone marrow cells into sub-lethally irradiated (5 Gy) *Rag1*^{-/-} mice, followed by reconstitution for at least 2 months. For treatment with tamoxifen, mice were injected intraperitoneally with tamoxifen (1 mg per mouse) in corn oil daily for 4 consecutive days and then analysed 7 days after the last injection. All mice were kept in a specific pathogen-free facility in the Animal Resource Center at St Jude Children's Research Hospital. Animal protocols were approved by the Institutional Animal Care and Use Committee of St Jude Children's Research Hospital.

Flow cytometry. For analysis of surface markers, cells were stained in PBS containing 2% (wt/vol) fetal bovine serum (FBS) with antibodies from eBioscience, unless otherwise noted. Foxp3 staining was performed as per the manufacturer's instruction (eBioscience). Intracellular staining of CTLA4 was performed together with Foxp3 using anti-CTLA4 antibody (UC10-4F10-11; BD Biosciences). BrdU and caspase-3 staining was performed as per the manufacturer's instruction (BD Biosciences). For detection of phosphorylated signalling proteins, lymphocytes were rested in complete medium for 1 h. They were fixed with Phosflow Lyse/Fix buffer, followed by permeabilization with Phosflow Perm buffer III (BD Biosciences) and staining with antibodies to S6 phosphorylated at Ser 235 and Ser 236 (D57.2.2E; Cell Signaling Technology), 4E-BP1 phosphorylated at Thr 37 and Thr 46 (236B4; Cell Signaling Technology), Akt phosphorylated at Ser 473 (M89-61; BD Biosciences) and Thr 308 (J1-223.371; BD Biosciences), and Erk phosphorylated at Thr 202 and Tyr 204 (20A; BD Biosciences). For staining mitochondria, lymphocytes were incubated for 30 min at 37 °C with 10 nM MitoTracker Deep Red (Life Technologies) or 20 nM TMRM (ImmunoChemistry Technologies) after staining surface markers. ROS were measured by incubation with 10 μ M 5-(and-6)-chloromethyl-2',7'-dichlorodihydrofluorescein diacetate, acetyl ester (CM-H2DCFDA; Life Technologies) or 2.5 μ M CellROX Deep Red (Life Technologies) after staining surface markers. Flow cytometry data were acquired on LSRII or LSR Fortessa (BD Biosciences) and analysed using Flowjo software (Tree Star).

Cell purification and culture. Lymphocytes were isolated from lymphoid organs (spleen and peripheral lymph nodes that included inguinal, axillary and cervical lymph nodes) and naive and T_{reg} cells were sorted on a MoFlow (Beckman-Coulter) or Reflection (i-Cyt). Sorted naive (CD4⁺CD62L⁺CD44^{lo}CD25⁻) and T_{reg} cells were used for *in vitro* culture in Click's medium (plus β -mercaptoethanol) supplemented with 10% (vol/vol) FBS and 1% (vol/vol) penicillin-streptomycin. T_{reg} cells were activated with plate-coated anti-CD3 (5 μ g ml⁻¹; 2C11; Bio X Cell), anti-CD28 (5 μ g ml⁻¹; 37.51; Bio X Cell) and human IL-2 (200 U ml⁻¹), in the presence or absence of simvastatin (2 μ M; EMD Millipore), or 25-hydroxycholesterol (0.5 μ g ml⁻¹; Sigma). Thymidine incorporation assays were performed by adding 0.2×10^5 T cells and 1×10^5 irradiated splenocytes together with soluble anti-CD3 (0.2 μ g ml⁻¹) in the presence or absence of human IL-2 (100 U ml⁻¹) for 72 h, followed by pulsing with [³H]-thymidine at 1 μ Ci per well for the last 12–16 h of culture. For labelling with CFSE and Celltrace violet (both from Life Technologies), T cells were incubated in Click's medium with 5% FBS and 4 μ M of either dye at 37 °C for 25 min, followed by extensive washes. *In vitro* T_{reg} suppression assays were performed as described⁷. Briefly, freshly sorted T_{eff} (CD4⁺CD45RB^{hi}CD25⁻; 5×10^4) and T_{reg} cells (at different ratios with T_{eff}) were cultured in 96-well plates along with anti-CD3 antibody and irradiated splenocytes for 72 h, followed by thymidine incorporation assays. For T_{reg} suppression assays using previously activated T_{reg} cells, T_{reg} cells were sorted from C57BL/6J mice and activated with anti-CD3, anti-CD28 and IL-2 for 3 days in the absence or presence of simvastatin

(2 μ M), 25-hydroxycholesterol (0.5 μ g ml⁻¹), 22(R)-hydroxycholesterol (5 μ M; Sigma), atorvastatin (5 μ M; EMD Millipore), lovastatin (5 μ M; EMD Millipore), mevalonate (100 μ M; Sigma), or certain combinations as specified; vehicle control did not have measurable effects (not shown). Viable T_{reg} cells were purified using Ficoll and incubated with freshly isolated T_{eff} cells for additional analysis as described earlier.

Colitis model. A total of 4×10^5 T_{eff} cells (CD4⁺CD45RB^{hi}CD25⁻) from CD45.1⁺ mice were mixed with 2×10^5 wild-type or raptor-deficient T_{reg} cells (from *Foxp3*^{cre}*Raptor*^{fl/fl} or *Cd4*^{cre}*Raptor*^{fl/fl} mice), and were transferred intraperitoneally into *Rag1*^{-/-} mice. Mice were assessed for clinical signs of colitis weekly and were analysed 8–10 weeks after transfer. Colons were fixed in 10% (vol/vol) neutral buffered formalin, and sectioned at 4 μ m thickness. Colon pathology was assigned scores by an experienced pathologist (P.V.) as described⁷. Lymphocytes were isolated from spleen and mesenteric lymph nodes and analysed by flow cytometry.

De novo lipogenesis, cholesterol measurement and bioenergetics assays. T_{reg} cells were activated with anti-CD3, anti-CD28 plus IL-2 for 20 h, and [1-¹⁴C]-acetic acid (Perkin Elmer; 4 μ Ci ml⁻¹) was added for an additional 4 h of culture. For labelling with [U-¹⁴C]-glucose (American Radiolabeled Chemicals), T_{reg} cells activated for 20 h were washed with PBS and incubated in glucose-free medium with dialysed FBS and [U-¹⁴C]-glucose (4 μ Ci ml⁻¹) for an additional 4 h of culture. After incubation, cells were collected, washed twice with PBS and lysed in 0.5% Triton X-100. Lipids were extracted by the addition of chloroform and methanol mixture (2:1 vol/vol) with vortexing, followed by the addition of water with vortexing. After centrifugation, the lipid-containing phase (at the bottom) was separated and ¹⁴C incorporation was measured using a Beckman LS6500 scintillation counter. Results are normalized to cell number. Cellular cholesterol level was measured using Amplex Red assay (Life Technologies). Briefly, cells were washed with PBS and lysed in the Amplex Red reaction buffer. After 15 min of incubation, cell lysates were centrifuged at maximum speed for 5 min. Fifty microlitres of supernatants were pipetted into a 96-well tissue culture plate and a 50 μ l aliquot of Amplex Red working solution was added to each well. The plate was incubated for 120 min at 37 °C, protected from light. Fluorescence was subsequently measured on a fluorescence microplate reader (Molecular Devices). A cholesterol standard curve was determined for each plate using a cholesterol reference standard. The bioenergetic activities of the ECAR and OCR pathways were measured using the Seahorse XF24-3 extracellular flux analyser as per the manufacturer's instructions (Seahorse Bioscience).

Gene expression profiling and gene-set enrichment analysis. RNA samples from freshly isolated T_{reg} cells from wild-type and *Cd4*^{cre}*Raptor*^{fl/fl} mice were analysed using the Affymetrix HT_MG-430_PM GeneTitan peg array, and expression signals were summarized using the RMA algorithm (Affymetrix Expression Console v.1.1). Gene-set enrichment analysis within canonical pathways was performed as described³².

RNA and immunoblot analysis. Real-time PCR analysis was done as described with primers and probe sets from Applied Biosystems⁷, or using the Power SYBR Green Master Mix from Life Technologies. Immunoblots were performed and quantified as described previously³¹, using the following antibodies: p-S6, p-Akt Ser 473, p-4E-BP1, p-Foxo1/3, raptor (all from Cell Signaling Technology) and β -actin (Sigma).

Statistical analysis. *P* values were calculated with Student's *t* test, Mann-Whitney test, or analysis of variance (GraphPad Prism) as specified in figure legends, with proper post-test analysis performed. Statistical analysis of mouse survival and respective *P* values were determined using the logrank test. **P* < 0.05; ***P* < 0.01; ****P* < 0.001.

- Yang, K., Neale, G., Green, D. R., He, W. & Chi, H. The tumor suppressor Tsc1 enforces quiescence of naive T cells to promote immune homeostasis and function. *Nature Immunol.* **12**, 888–897 (2011).
- Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

An ultra-lightweight design for imperceptible plastic electronics

Martin Kaltenbrunner^{1,2,3}, Tsuyoshi Sekitani^{1,2}, Jonathan Reeder^{1†}, Tomoyuki Yokota¹, Kazunori Kuribara¹, Takeyoshi Tokuhara¹, Michael Drack³, Reinhard Schwödau³, Ingrid Graz³, Simona Bauer-Gogonea³, Siegfried Bauer³ & Takao Someya^{1,2}

Electronic devices have advanced from their heavy, bulky origins to become smart, mobile appliances. Nevertheless, they remain rigid, which precludes their intimate integration into everyday life. Flexible, textile and stretchable electronics are emerging research areas and may yield mainstream technologies^{1–3}. Rollable and unbreakable backplanes with amorphous silicon field-effect transistors on steel substrates only 3 μm thick have been demonstrated⁴. On polymer substrates, bending radii of 0.1 mm have been achieved in flexible electronic devices^{5–7}. Concurrently, the need for compliant electronics that can not only be flexed but also conform to three-dimensional shapes has emerged³. Approaches include the transfer of ultrathin polyimide layers encapsulating silicon CMOS circuits onto pre-stretched elastomers⁸, the use of conductive elastomers integrated with organic field-effect transistors (OFETs) on polyimide islands⁹, and fabrication of OFETs and gold interconnects on elastic substrates¹⁰ to realize pressure, temperature and optical sensors^{11–14}. Here we present a platform that makes electronics both virtually unbreakable⁴ and imperceptible. Fabricated directly on ultrathin (1 μm) polymer foils, our electronic circuits are light (3 g m^{–2}) and ultraflexible and conform to their ambient, dynamic environment. Organic transistors with an ultra-dense oxide gate dielectric a few nanometres thick formed at room temperature enable sophisticated large-area electronic foils with unprecedented mechanical and environmental stability: they withstand repeated bending to radii of 5 μm and less, can be crumpled like paper, accommodate stretching up to 230% on prestrained elastomers, and can be operated at high temperatures and in aqueous environments. Because manufacturing costs of organic electronics are potentially low, imperceptible electronic foils may be as common in the future as plastic wrap is today. Applications include matrix-addressed tactile sensor foils for health care and monitoring, thin-film heaters, temperature and infrared sensors, displays¹⁵, and organic solar cells¹⁶.

An example of an ultrathin active-matrix array with resistive tactile sensors is shown in Fig. 1a. The electronic circuit foil weighs so little that it floats through air much like a feather (Fig. 1b and Supplementary Movie 1). At only 3 g m^{–2}, our sensor foil is 60-fold lighter than conventional polyimide substrates (125 μm thick, 180 g m^{–2}), 27-fold lighter than office paper (130 μm thick, 80 g m^{–2}) and still eightfold lighter than a steel foil 3 μm thick (24.2 g m^{–2}). The unprecedented mechanical properties of the foil are illustrated in Fig. 1c, in which it survives being crumpled like a piece of paper. These imperceptible ultrathin foils form an intimate contact with arbitrary shapes (including human skin) without imposing constraints either mechanically or by weight. In the ultrathin array in Fig. 1, a polyethylene naphthalate (PEN) foil 1 μm thick was used as substrate; the circuits were formed with a hybrid anodic aluminium oxide^{17,18} and a phosphonic acid self-assembled monolayer¹⁹ (SAM) gate dielectric, dinaphtho[2,3-b:2',3'-f]thieno[3,2-b]thiophene (DNTT) as air-stable organic semiconductor²⁰, and a parylene interlayer separating the active-matrix backplane from

the sensing layer. The fabrication of the device consists exclusively of process steps commonly used in the semiconductor industry, such as vacuum evaporation (metals and organic semiconductor) and chemical vapour deposition (parylene). Potentiostatic anodization was used to form ultra-dense and durable high-*k* dielectric layers a few nanometres thick for the organic transistors¹⁷. The self-healing nature of the anodic oxide enables us to fabricate complex devices as thin as 100 nm on industrial-scale plastic substrates with considerable (88 nm root mean square) surface roughness, while maintaining high performance (see Methods and Supplementary Information for details). Typically, all 12 \times 12 sensor cells on an active-matrix array are functional.

Figure 2 shows potential applications of our technology as an infrared bolometer sensor (Fig. 2a, b) and in heat management, where high electrical power is effectively dissipated to a heat sink (Fig. 2c–j). For these applications, a thin gold metal film 100 nm thick was evaporated onto the 1- μm PEN foil. The transient temperature signals in Fig. 2b show temperature equilibration within 9 μs after excitation with a 10-ns laser pulse from an Nd:YAG laser at 1,064 nm wavelength. The bolometer has a small thermal mass of $C = 8 \mu\text{J K}^{-1}$, comparable to micromachined devices^{21,22}. Fast temperature equilibration allows efficient power dissipation from the device to an underlying substrate, thus enabling efficient heat management strategies in flexible and stretchable electronics, an emerging topic^{23,24}. As depicted in the setup photograph and in the series of infrared thermal images in Fig. 2c–j, the electronic foil dissipates an electrical power of 0.35 W to a Teflon substrate and a 20-fold higher power to a heat sink substrate with current densities comparable to those found in integrated circuits²⁵. An example of its use as a clinging temperature sensor, wrapped around a capillary of a high-performance liquid chromatography pipe, is shown in Supplementary Fig. 1. Localized heating, as in the example of a thin metal-foil sensor for temperature sensing and heating on a human finger, is demonstrated in Supplementary Fig. 2a, b.

We added complexity to the devices by fabricating thin-film transistors for actively controlled sensors and actuator networks. Aluminium serves as the gate electrode, potentiostatically anodized aluminium oxide is the gate dielectric, an n-tetradecyl phosphonic acid SAM provides the interface to the air-stable high-mobility p-type organic semiconductor DNTT. Gold source and drain contacts complete the device (Fig. 3a). An optical micrograph of a single transistor on the 1.2- μm PEN substrate (channel length 40 μm , width 500 μm) is shown in Fig. 3b. A detailed analysis of the surface morphology by atomic force microscopy is provided in Supplementary Fig. 3. For mechanical testing devices were encapsulated with a parylene layer 0.8–1 μm thick that places the thin-film transistor near the neutral plane position^{5,6,8} (see Methods). A cross-sectional scanning electron microscopy (SEM) image of an encapsulated device with a total thickness of 2 μm is shown in Fig. 3c. The layer structure of the 100-nm OFET is revealed in the scanning transmission electron microscope (STEM) cross-sectional image (Fig. 3d). The amorphous nature of the aluminium oxide layer

¹The University of Tokyo, Electrical and Electronic Engineering and Information Systems, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. ²Exploratory Research for Advanced Technology (ERATO), Japan Science and Technology Agency (JST), 2-11-16, Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan. ³Johannes Kepler University, Soft Matter Physics, Altenbergerstrasse 69, 4040 Linz, Austria. [†]Present address: The University of Texas at Dallas, Department of Materials Science and Engineering, 800 West Campbell Road, Richardson, Texas 75080-3021, USA.

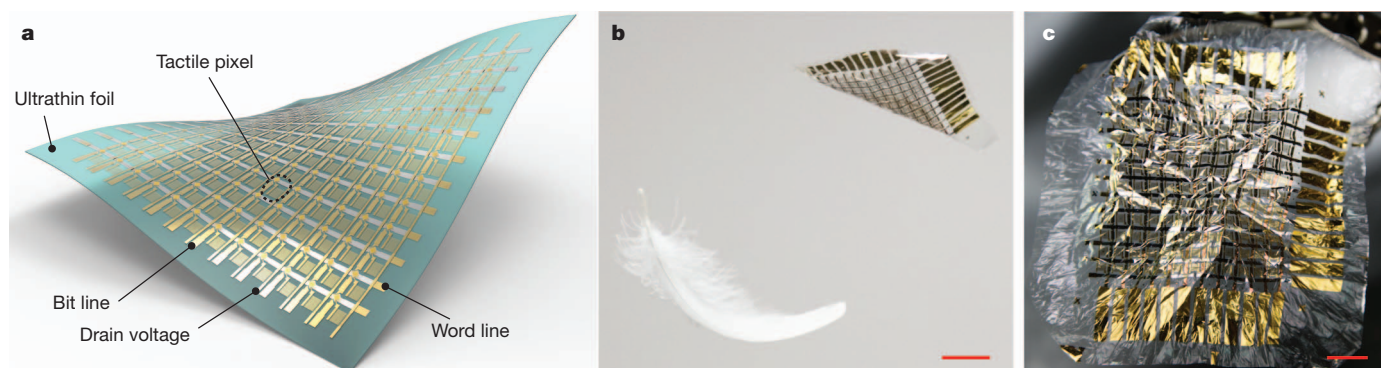


Figure 1 | Imperceptible electronic foil. **a**, Illustration of a thin large-area active-matrix sensor with 12×12 tactile pixels. **b**, Ultrathin plastic electronic foils are extremely lightweight (3 g m^{-2}); they float to the ground more slowly

than a feather and are therefore virtually unbreakable. Scale bar, 2 cm. **c**, At only $2 \mu\text{m}$ thickness, our devices are ultraflexible and can be crumpled like a sheet of paper. Scale bar, 1 cm.

is confirmed by the absence of crystalline regions in the bright-field STEM image in Supplementary Fig. 4a. The elemental composition of the individual layers was analysed by means of an energy-dispersive X-ray system (Supplementary Fig. 4b, c), which confirmed the formation of a SAM layer. We then compared our anodic aluminium oxide hybrid dielectric with conventional aluminium oxide prepared by treatment with oxygen plasma (Supplementary Fig. 5). In short, potentiostatically anodized alumina shows higher mechanical durability than plasma-formed alumina, and withstands large bending strains at relevant operating voltages. This, together with the ultrathin architecture, allows our transistors to survive crumpling like a sheet of paper without performance degradation (Supplementary Fig. 6). Transfer and output characteristics of an OFET with a saturation mobility of $3 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ are plotted in Fig. 3e, f; Supplementary Fig. 7 characterizes devices with 9 nm gate oxide. The devices have a performance equal to that of reference devices fabricated on rigid, smooth Si/SiO₂

substrates (Supplementary Fig. 8), and are on a par with the best-performing OFETs reported so far²⁶.

To demonstrate the potential of our technology, we fabricated a 12×12 active-matrix array with integrated resistive tactile sensors. The diagram in Fig. 4 shows the design of a tactile pixel. The control transistors were fabricated on the $1.2\text{-}\mu\text{m}$ substrate with a high aspect ratio (channel width-to-length ratio $12,000 \mu\text{m}/20 \mu\text{m}$), which makes the active-matrix backplane suitable for use in high-power applications, for instance as OLED pixel drivers. After the formation of an insulating interlayer of 800-nm parylene, interdigitated gold electrodes were deposited as resistive touch sensors. The electrical connection between the touch sensors and the switching matrix backplane is provided by gold evaporated through high-resolution vias drilled with a green laser marker system. The rough surface of the parylene interlayer drastically improves the adhesion of the gold sensor layer, providing a robust and wear-resistant tactile sensing pixel (Supplementary

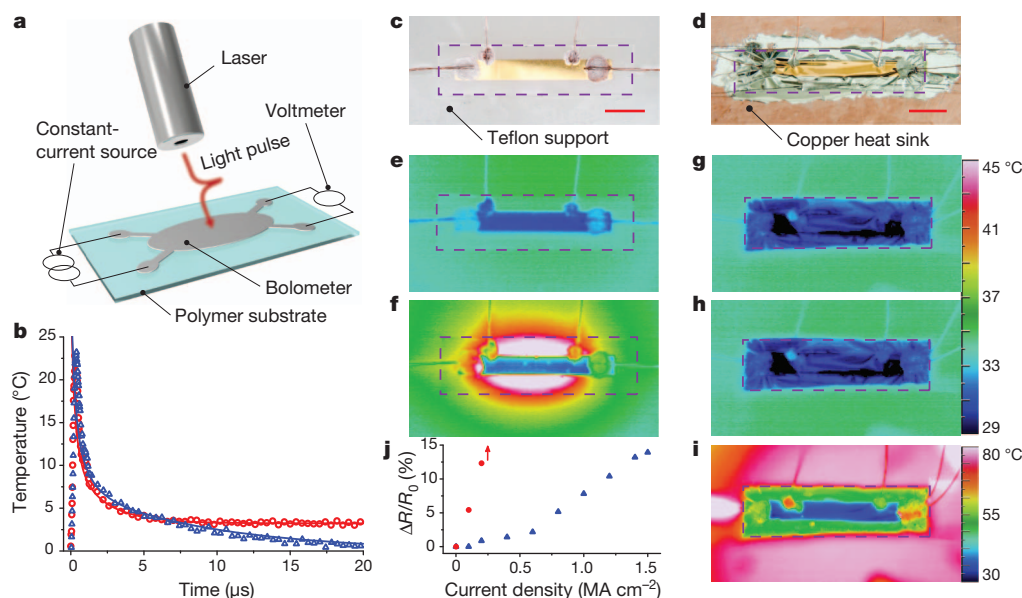


Figure 2 | Thin-film infrared sensor and heat management. **a**, Diagram of a thin-film infrared bolometer sensor and experimental setup to measure the transient temperature signal on excitation by a 1,064-nm Nd:YAG laser pulse. **b**, Temperature change after thermal pulse excitation of a free-standing bolometer (red circles) and a bolometer coupled to a heat sink (blue triangles). Heat is transferred to the heat sink very rapidly (on a timescale of only $4\tau = 9 \mu\text{s}$). Solid lines are fits according to equations (1) and (2) in Supplementary Information. **c–j**, Heat management in thin-film electronics. Metallic conductors (100-nm Au) on a $1.2\text{-}\mu\text{m}$ PEN foil are placed on a Teflon support with low thermal conductivity (**c**) and on a copper block serving as heat

sink (**d**). Scale bars, 2 mm. The excellent heat transport properties of such a thin-foil substrate allow these devices to handle the current densities used in integrated circuits and to dissipate a significant amount of power when thermally connected to a heat sink. Thermography images visualize dissipated heat on Teflon at zero current (**e**) and 0.2 A (**f**) and the copper support at zero current (**g**), 0.2 A (**h**) and 1.5 A (**i**). **j**, Resistance change $\Delta R/R_0$ of a conductor on Teflon support (red circles) and in thermal contact with a copper heat sink (blue triangles). While Teflon supported conductors fail at a current density of 0.2 MA cm^{-2} , a current density tenfold higher is sustained when the device is placed on a heat sink.

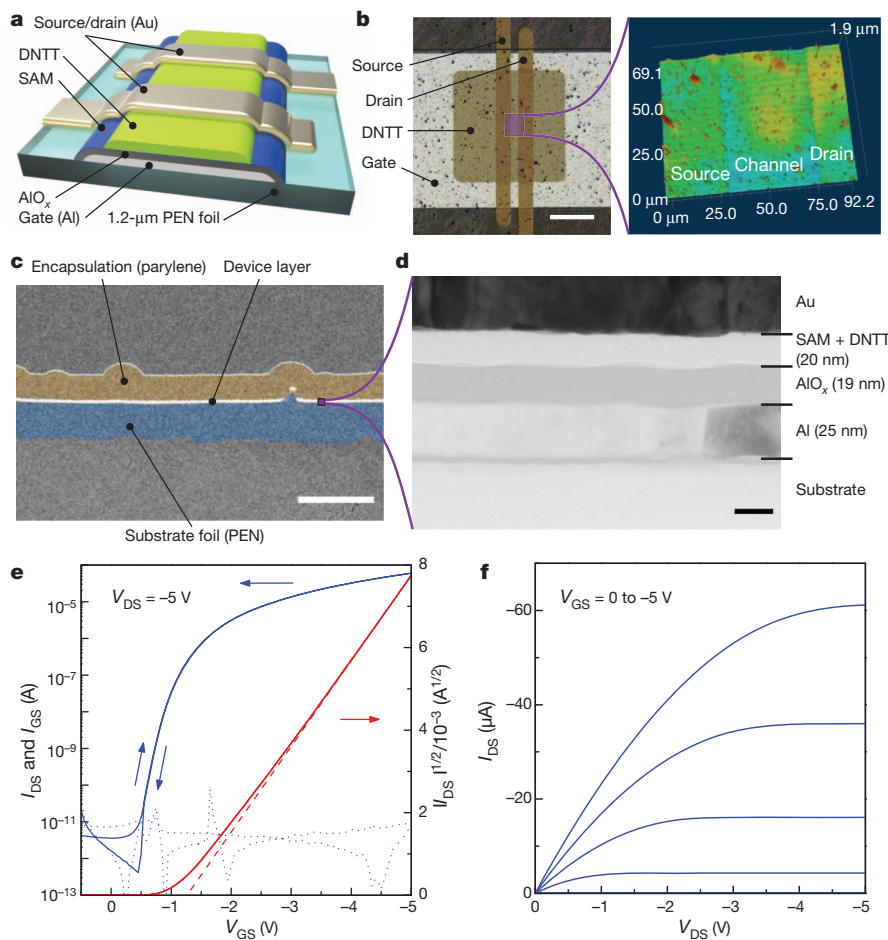


Figure 3 | Thin-film transistors on 1.2- μm PEN foils. **a**, Cross-sectional diagram of an ultrathin OFET. Aluminium serves as gate contact and a mechanically durable and dense hybrid dielectric is formed from potentiostatically anodized alumina and a phosphonic acid SAM. The organic semiconductor DNTT ensures air-stable operation, and 50 nm thick Au serves as the source and drain. The complete device including the PEN substrate is 1.3 μm thick. **b**, Optical micrograph of a single transistor on the 1.2- μm PEN substrate with a channel length of 40 μm and width of 500 μm . Scale bar, 200 μm . The magnified three-dimensional view of the channel region highlights the topology of the transistor on the unconventionally rough substrate. **c**, Cross-sectional SEM image of the ultrathin transistor encapsulated with 800-nm parylene diX-SR to protect against mechanical damage and place the transistor near the neutral strain position. The total thickness of this device is about 2 μm . Scale bar, 2 μm . **d**, High-resolution cross-sectional TEM image showing the layer structure of the 100-nm OFET. The amorphous aluminium oxide layer 19 nm thick ensures low leakage currents even on rough substrates and is mechanically reliable, while maintaining low operating voltages. An n-tetradecyl phosphonic acid SAM about 2 nm thick forms the interface to a DNTT layer roughly 18 nm thick. Scale bar, 20 nm. **e**, Transfer characteristics of an ultrathin, flexible OFET. Plotted are drain-source current (I_{DS}) and gate-source current (I_{GS}) as a function of gate-source voltage (V_{GS}) at a drain-source voltage (V_{DS}) of -5 V . The saturation mobility is $3\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$, the on/off ratio exceeds seven orders of magnitude and the gate leakage current (I_{GS}) is less than 10 pA. **f**, Corresponding output characteristics. Plotted are I_{DS} as a function of V_{DS} for a gate-source voltage from 0 V to -5 V .

Fig. 9). Figure 4b shows a top-view photograph of the completed $8\text{ cm} \times 8\text{ cm}$ ultrathin tactile sensor array; a magnified view of four pixels is shown in the inset. A detailed view of a single tactile pixel is shown in Fig. 4c, together with a single backplane transistor without the sensor layer and a magnified view of the transistor's active region. A circuit diagram of the resistive sensor pixel illustrates the operating principle of the sensor cell. Figure 4d plots the transfer characteristic of a transistor in the active-matrix configuration (without the sensor layer). Our transistors maintain saturation mobilities of up to $1.6\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$ when operated at only 3 V source drain and gate voltage and an on/off-current ratio of eight orders of magnitude. The ultra-dense hybrid dielectric layer ensures low gate leakage currents of less than 100 pA even for transistors with a high channel width-to-length ratio. Using a high width-to-length ratio enables output currents of more than 0.5 mA at 3 V, which is sufficient for OLED display drivers and power amplifiers. The average mobility across all 144 transistors on the tactile sensor foil is $0.88 \pm 0.13\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$, as the mobility distribution plotted in Fig. 4e shows. A critical design parameter for active-matrix-based sensor arrays is the on/off ratio of the control transistors, which must be large enough to allow unambiguous distinction between on and off pixels. The on/off ratio for all of the 144 pixels of our ultrathin tactile sensor exceeds 10^7 (on average $(2.57 \pm 0.7) \times 10^7$; see distribution in Supplementary Fig. 10f). Threshold voltage, sub-threshold swing and on-current and off-current distributions are shown in Supplementary Fig. 10. When connected, the drain current of a tactile sensing pixel depends on the surface conductivity of the sensed object. Figure 4f shows the transfer curve for a single pixel when the interdigitated resistive sensor is in contact with various objects. Highly

conducting surfaces are readily detected, and even non-conducting objects can be perceived when an anisotropically conducting pressure-sensitive rubber is laminated to the tactile sensing array (a soft touch with 1 kPa pressure is readily registered). Such tactile sensing foils can be used as human-machine interfaces, because the resistivity of human skin is well suited to activating a pixel with a gentle touch. More details and high-resolution images of the individual fabrication steps are given in Supplementary Fig. 11. To demonstrate the functionality as a tactile sensor, we recorded the position of a metallic ring when placed on the sensing sheet. Figure 4g shows a top-view photograph of the ring and the corresponding signal map. The circular pattern is read out with high precision, and a difference of more than six orders of magnitude in the on-current between contacted and non-contacted pixels is observed. Such ultrathin large-area sensing sheets readily conform to three-dimensional surfaces, allowing unprecedented applications: displays, sensors, thermal imaging in sports and recreation, medical treatment and prosthetics; electronic skin; robotics; and consumer electronics. The example of a tactile sensor conforming to a model of the upper human jaw (Fig. 4h) illustrates potential applications in health care and monitoring, in which imperceptible sensing and actuating systems are required to minimize patient discomfort. In biomedical applications, including *in vivo* sensing of biosignals, ultrathin electronic foils must operate at elevated temperatures and in aqueous environments. At only 2 μm thickness (including parylene encapsulation), our transistors can withstand thermal annealing at up to 170°C (Supplementary Fig. 12) and operate continuously for more than two weeks fully immersed in saline solution without significant degradation in performance (Supplementary Fig. 13).

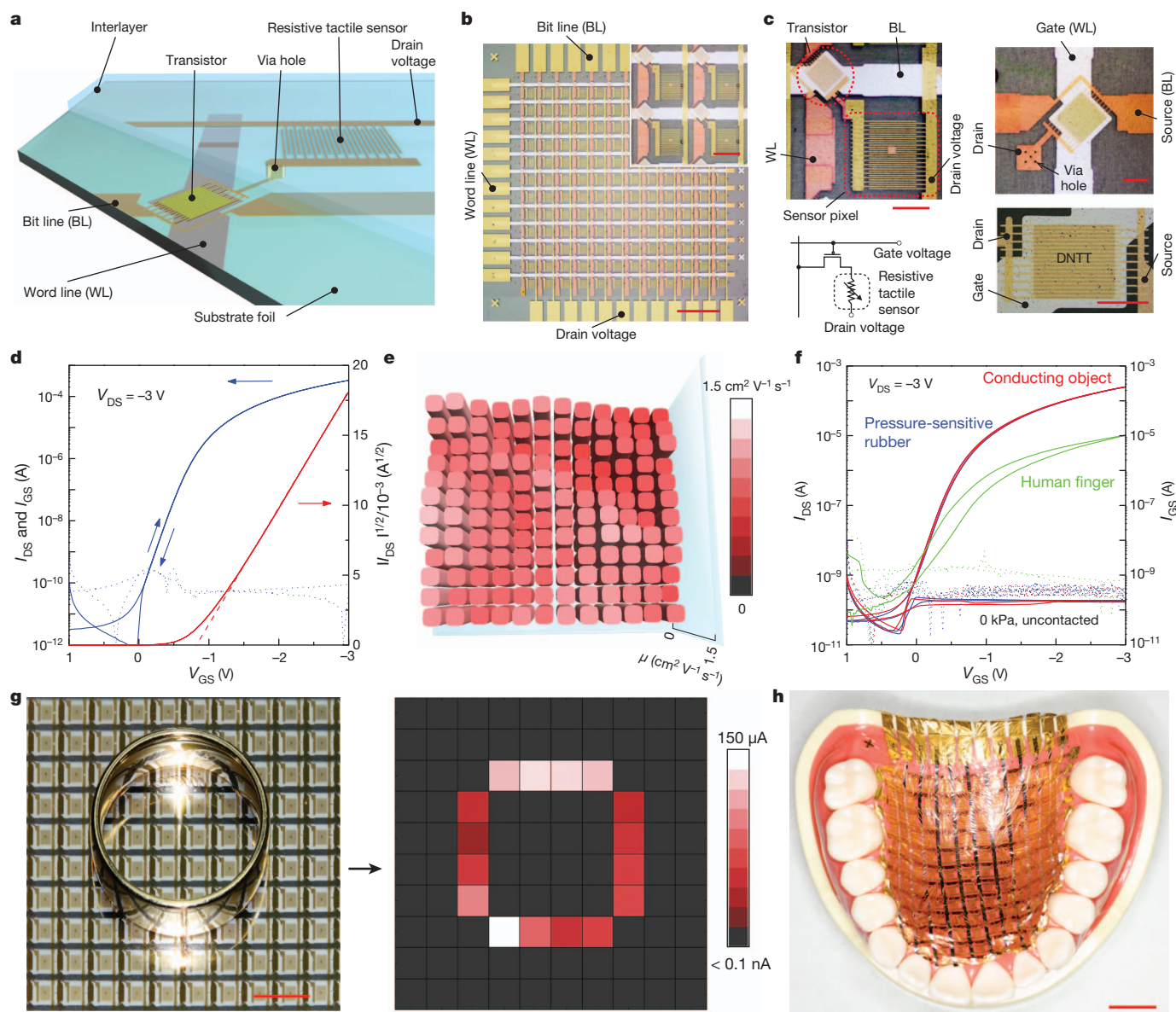


Figure 4 | Active-matrix tactile sensing foil. **a**, Design of a tactile pixel comprising a switching transistor and a resistive touch sensor. The control transistor is fabricated directly on the 1.2- μm substrate. An interlayer of 800-nm parylene insulates the switching matrix backplane from the sensing layer. The electrical connection to the transistors is provided through high-resolution vias. Interdigitated gold electrodes with a line pitch of 30 μm serve as touch sensors. **b**, Top-view photograph of the completed 8 cm \times 8 cm ultrathin tactile sensor array with 12 \times 12 pixels. Scale bar, 1 cm. A magnified view of four pixels is shown in the inset. (Scale bar: 2 mm) **c**, Top: detailed view of an individual tactile pixel together with a single backplane transistor (left) and without the sensor layer (right). Scale bars, 1 mm. Bottom: circuit diagram of the resistive sensor pixel (left) and magnified view of the transistor's active region (right).

To demonstrate the stretch compatibility of our lightweight electronic foils, we laminated devices to a pre-stretched elastomeric film. When the pre-stretched elastomer is allowed to relax, wrinkles form in the active transistor layer, which enables stretching of the device in the direction of the initial deformation of the elastomer^{3,27–30}. The electronic circuits were fabricated close to the neutral plane of the foil stack, which permits bending to a radius of as little as 5 μm with marginal change in electrical performance. Figure 5a illustrates how stretchability of ultrathin electronics is achieved: after fabrication of the devices on the carrier substrate, the electronic film is peeled off the support stack and laminated to a pre-stretched elastomer (3M VHB)

Scale bar, 500 μm . **d**, Transfer characteristic of a transistor in the matrix configuration. The saturation mobility is 1.6 $\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$ at 3 V source-drain and gate voltage, on/off-current ratio is 10^8 . The ultra-dense dielectric layer ensures low gate leakage currents (less than 100 pA), enabling output currents of more than 0.5 mA at 3 V. **e**, Typical mobility distribution of all 144 transistors on a completed tactile sensor. The average mobility is $0.88 \pm 0.13 \text{ cm}^2 \text{V}^{-1} \text{s}^{-1}$. **f**, Transfer curve for a single pixel when the interdigitated resistive sensor is touched by various objects. **g**, Top-view photograph (left) and corresponding drain current of a metallic ring placed on the sensing sheet (right). A difference of more than six orders of magnitude is observed in drain current between contacted and non-contacted pixels. Scale bar, 1 cm. **h**, Tactile sensor sheet tightly conforming to a model of the human upper jaw. Scale bar, 1 cm.

by simple picking and placing. When the prestrain in the elastomer is relaxed, adhesion between the elastomer and the thin electronic foil causes the strain in the elastomer to transfer to the thin foil, forming a network of out-of-plane wrinkles in the device that accommodates subsequent tensile strains. Figure 5b shows a picture sequence of this formation of folds when the elastomer is relaxed. The wrinkle topology was characterized by both three-dimensional optical microscopy and SEM (Fig. 5c). To determine the minimum bending radii of the multiple folds that form in the active area of the transistors, we analysed cross-sectional SEM images of a transistor undergoing 233% tensile strain. Figure 5d reveals multiple folds with bending radii of less than

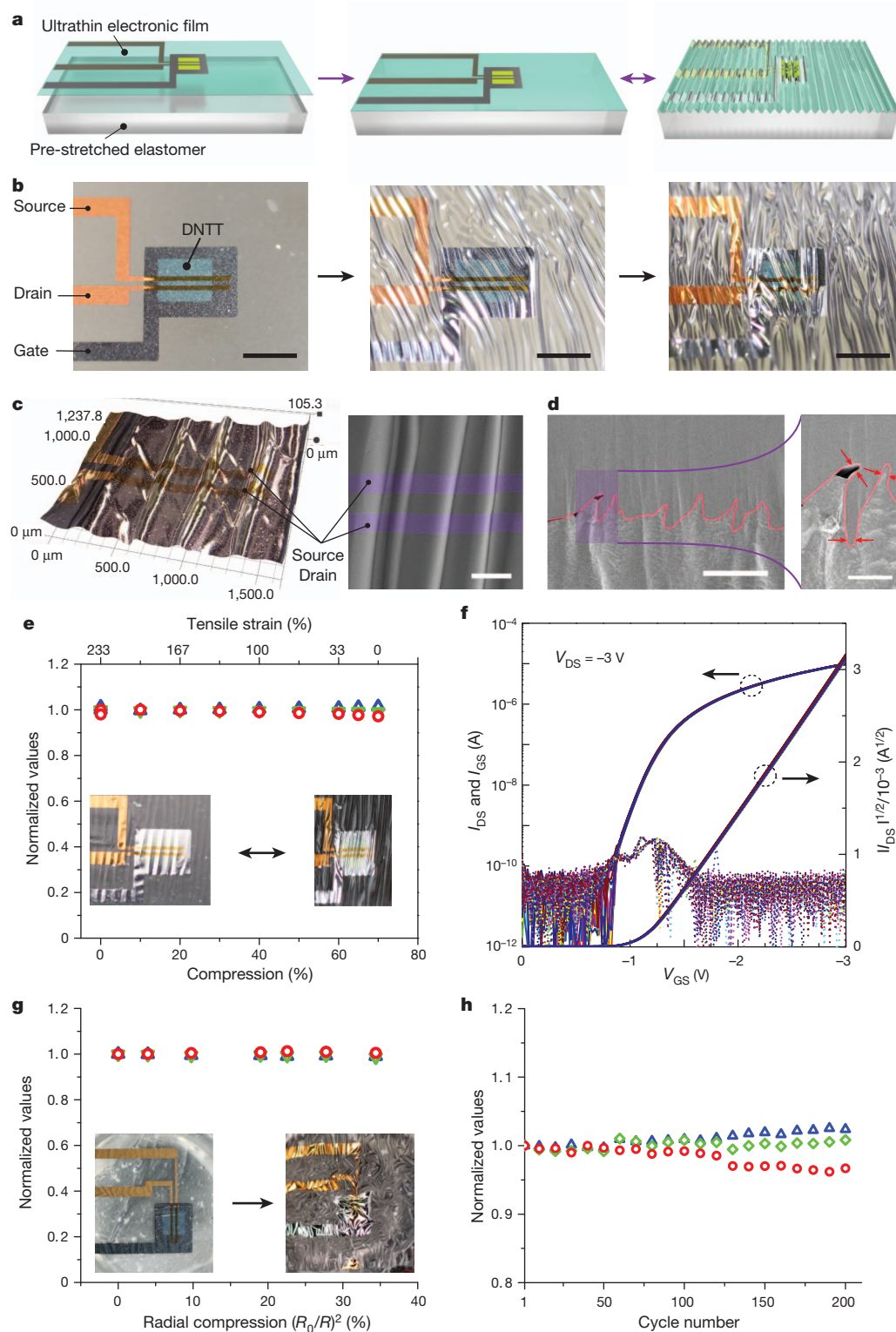


Figure 5 | Stretch-compatible ultraflexible transistors.

a, Illustration of stretchable ultrathin electronics. After fabrication, devices are peeled off the support stack and laminated to a pre-stretched elastomer (3M VHB) by simple picking and placing. When the prestrain in the elastomer is relaxed, the ultrathin electronic foil forms a network of out-of-plane wrinkles that accommodate subsequent tensile strains. **b**, Picture sequence of the formation of folds in an ultraflexible transistor when the elastomer is relaxed. Scale bars, 1 mm. **c**, Wrinkle topology characterized by a three-dimensional optical microscope (left) and by SEM (right). The out-of-plane folds are about 100 μm high. Scale bar, 200 μm . **d**, Cross-sectional SEM image of a transistor undergoing 70% compressive strain. The magnified view reveals multiple folds with bending radii of less than 5 μm within the active device area. Scale bars, 100 μm (left); 20 μm (right). **e**, On-current (red circles), mobility (green diamonds) and threshold voltage (blue triangles) as functions of compressive and tensile strains. No irreversible degradation was observed during the experiment. Inset: a transistor in relaxed (left) and compressed (right) states. **f**, Transfer curves corresponding to **e**, demonstrating the mechanical stability of the hybrid gate dielectric: no increase in gate leakage current was observed. **g**, Biaxially stretchable transistor. When laminated to an equi-biaxially prestretched elastomer, these transistors withstand a 35% decrease in area before failure. The inset shows the transistor in relaxed (left) and compressed (right) states. **h**, Mechanical durability during repeated compression and re-stretching to 100% tensile strain. Plot of on-current (red circles), mobility (green diamonds) and threshold voltage (blue triangles) against stretch cycles. Even after 200 full cycles, less than 4% change in device parameters was observed.

5 μm within the active device area. Despite these extreme bending conditions, the transistors not only remain fully functional, but, as a result of the neutral-plane fabrication, there is virtually no performance degradation even at 233% tensile strain. Figure 5e shows on-current, mobility and threshold voltage as functions of compressive and tensile strains. The maximum compression is determined by the pre-strain of the elastomer and the packing density of the out-of-plane folds. No irreversible degradation was observed during the experiment. The corresponding transfer curves (Fig. 5f) demonstrate the

superior mechanical stability of our hybrid gate dielectric: no increase in gate leakage current was observed. In a biaxially pre-stretched elastomer, a random network of wrinkles forms on relaxation. These devices remain fully functional up to a 35% decrease in area (Fig. 5g); beyond this value, the bending strain on the active device becomes too great and dielectric breakdown occurs. Because no measures were taken to control the formation of wrinkles, we expect selective bonding of the electronic foil to the pre-stained elastomer to further improve performance in this area. Finally, we explored the mechanical durability of uniaxially

stretchable devices by repeated compression and re-stretching to 100% tensile strain. An image sequence taken during a stretching cycle is shown in Supplementary Fig. 14 and Supplementary Movie 2. Even after 200 full cycles the change in device parameters amounts to less than 4% (Fig. 5h), which—estimated conservatively—suggests a t_{80} lifetime of more than 1,000 full cycles (t_{80} is the number of cycles until device parameters degrade to 80% of their initial values). More information on the stretching test setup and additional experimental results is given in Supplementary Figs 15–17.

The technology described here is based on common organic electronic manufacturing methods, and enables the fabrication of imperceptible electronic systems containing integrated circuits, sensors and power supplies. At only 2 μm total thickness, these electronic devices are extremely lightweight and minimize electronic waste; like plastic wrap, they are easily applied to curvilinear and dynamic surfaces and can therefore be used in consumer and mobile electronic appliances, architectural design, robotics, emergency response, sports, healthcare and biomedical systems.

METHODS SUMMARY

PET sheets (thickness 125 μm) covered with a thin polydimethylsiloxane layer served as supporting substrates to ease handling of the 1.2- μm PEN foil for laboratory-scale device fabrication. The PEN film (obtained from Teijin Dupont Films, Japan) adheres to the supporting substrate only by weak van der Waals forces, which readily enables all subsequent processing steps (vacuum evaporation, anodization and chemical vapour deposition). After fabrication, the devices can be peeled off the supporting stack without causing damage. Critical for low-voltage, high-performance OFETs on very thin, ultra-rough foils is a dense, nanometre-thick gate dielectric layer that can be processed at room temperature. Potentiostatic anodization was found to form self-healing nanoscale barrier oxides disregarding the substrate roughness. The technique is readily scalable and complies with industrial ultrathin polymer substrates.

The position b of the neutral mechanical plane (zero-strain position) of a multi-layer stack with the n th layer on top of the first layer on the bottom is given by⁸

$$b = \frac{\sum_{i=1}^n E_i t_i \left[\sum_{j=1}^i t_j - \frac{t_i}{2} \right]}{\sum_{i=1}^n E_i t_i} \quad (1)$$

where E_i and t_i denote the Young's moduli and thicknesses of the individual layers. For the ultrathin OFETs ($n = 6$; 1.2 μm PEN/25 nm Al/19 nm anodic AlO_x /20 nm DNTT/40 nm Au/0.8 to 1 μm parylene diX-SR) the elastic moduli are $E_{\text{PEN}} = 4$ GPa, $E_{\text{Al}} = 69$ GPa, $E_{\text{AlO}_x} = 130$ GPa, $E_{\text{DNTT}} = 2.3$ GPa, $E_{\text{Au}} = 79$ GPa and $E_{\text{diX-SR}} = 4$ GPa. This places the neutral mechanical plane at a distance b of 1.15–1.2 μm from the bottom surface of the PEN substrate, near the position of the active OFET structure. Future theoretical work should focus not only on wrinkle frequencies and amplitudes but also on instabilities in wrinkle geometry at large stretch, as well as possible delamination processes of the ultrathin films on elastomeric substrates and heat transport in such multilayer systems.

Full Methods and any associated references are available in the online version of the paper.

Received 28 February; accepted 14 May 2013.

- Wong, W. S. & Salleo, A. (eds) *Flexible Electronics: Materials and Applications* (Springer, 2010).
- Cherenack, K. & van Pieterse, L. Smart textiles: challenges and opportunities. *J. Appl. Phys.* **112**, 091301 (2012).
- Wagner, S. & Bauer, S. Materials for stretchable electronics. *MRS Bull.* **37** (special issue), 207–213 (2012).
- Ma, E. Y. & Wagner, S. Amorphous silicon transistors on ultrathin steel foil substrates. *Appl. Phys. Lett.* **74**, 2661–2662 (1999).
- Sekitani, T. *et al.* Ultraflexible organic field-effect transistors embedded at a neutral strain position. *Appl. Phys. Lett.* **87**, 173502 (2005).
- Sekitani, T., Zschieschang, U., Klauk, H. & Someya, T. Flexible organic transistors and circuits with extreme bending stability. *Nature Mater.* **9**, 1015–1022 (2010).

- Shahrjerdi, D. & Bedell, S. W. Extremely flexible nanoscale ultrathin body silicon integrated circuits on plastic. *Nano Lett.* **13**, 315–320 (2013).
- Kim, D. H. *et al.* Stretchable and foldable silicon integrated circuits. *Science* **320**, 507–511 (2008).
- Sekitani, T. *et al.* A rubberlike stretchable active matrix using elastic conductors. *Science* **321**, 1468–1472 (2008).
- Graz, I. M., Cotton, D. P. J., Robinson, A. & Lacour, S. P. Silicone substrate with *in situ* strain relief for stretchable thin-film transistors. *Appl. Phys. Lett.* **98**, 124101 (2011).
- Someya, T. *et al.* Conformable, flexible, large-area networks of pressure and thermal sensors with organic transistor active matrixes. *Proc. Natl Acad. Sci. USA* **102**, 12321–12325 (2005).
- Kim, D. H. *et al.* Epidermal electronics. *Science* **333**, 838–843 (2011).
- Sokolov, A. N., Tee, B. C. K., Bettinger, C. J., Tok, J. B. H. & Bao, Z. Chemical and engineering approaches to enable organic field-effect transistors for electronic skin applications. *Acc. Chem. Res.* **45**, 361–371 (2012).
- Lipomi, D. J. *et al.* Skin-like pressure and strain sensors based on transparent elastic films of carbon nanotubes. *Nature Nanotechnol.* **6**, 788–792 (2011).
- White, M. S. *et al.* Ultrathin, highly flexible and stretchable PLEDs. *Nature Photonics*. (in the press).
- Kaltenbrunner, M. *et al.* Ultrathin and lightweight organic solar cells with high flexibility. *Nature Commun.* **3**, 770–777 (2012).
- Mardare, A. I., Kaltenbrunner, M., Sariciftci, N. S., Bauer, S. & Hassel, A. W. Ultra-thin anodic alumina capacitor films for plastic electronics. *Phys. Status Solidi, A Appl. Res.* **209**, 813–818 (2012).
- Lohregel, M. M. Thin anodic oxide layers on aluminum and other valve metals—high-field regime. *Mater. Sci. Eng. Rep.* **11**, 243–294 (1993).
- Klauk, H., Zschieschang, U., Pfaff, J. & Halik, M. Ultralow-power organic complementary circuits. *Nature* **445**, 745–748 (2007).
- Yamamoto, T. & Takimiya, K. Facile synthesis of highly pi-extended heteroarenes, dinaphtho[2,3-b:2',3'-f]chalcogenopheno[3,2-b]chalcogenophenes, and their application to field-effect transistors. *J. Am. Chem. Soc.* **129**, 2224–2225 (2007).
- de Almeida, L. A. L. *et al.* Modeling and performance of vanadium-oxide transition edge microbolometers. *Appl. Phys. Lett.* **85**, 3605–3607 (2004).
- Wang, B., Lai, J., Li, H., Hu, H. & Chen, S. Nanostructured vanadium oxide thin film with high TCR at room temperature for microbolometer. *Infrared Phys. Technol.* **57C**, 8–13 (2013).
- Oh, D.-W., Kim, S., Rogers, J. A., Cahill, D. G. & Sinha, S. Interfacial thermal conductance of transfer-printed metal films. *Adv. Mater.* **23**, 5028–5033 (2011).
- Gonzalez, M. *et al.* in *11th International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE 2010)* (eds Ernst, L. J. *et al.*) 1–7 (IEEE, 2010).
- Tu, K. N. Recent advances on electromigration in very-large-scale-integration of interconnects. *J. Appl. Phys.* **94**, 5451–5473 (2003).
- Zschieschang, U. *et al.* Flexible low-voltage organic thin-film transistors and circuits based on C10-DNTT. *J. Mater. Chem.* **22**, 4273–4277 (2012).
- Khang, D., Jiang, H., Huang, Y. & Rogers, J. A stretchable form of single-crystal silicon for high-performance electronics on rubber substrates. *Science* **311**, 208–212 (2006).
- Suo, Z., Ma, E., Gleskova, H. & Wagner, S. Mechanics of rollable and foldable film-on-foil electronics. *Appl. Phys. Lett.* **74**, 1177–1179 (1999).
- Rogers, J. A., Someya, T. & Huang, Y. Materials and mechanics for stretchable electronics. *Science* **327**, 1603–1607 (2010).
- Zang, J. *et al.* Multifunctionality and control of the crumpling and unfolding of large-area graphene. *Nature Mater.* **12**, 1–5 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank I. Abfalter for discussions. This work was supported by the JST Someya Bio-Harmonized ERATO grant and the ERC Advanced Investigators Grant 'Soft-Map' of S. Bauer. M.K. acknowledges financial support from the Wilhelm Macke Foundation and the mobility programme of the Johannes Kepler University Linz (KIP).

Author Contributions M.K. and T.Se. designed, fabricated and characterized the transistors and active matrices. J.R., T.Y., K.K. and T.T. performed device fabrication and characterization. S.B.-G., R.S. and S.B. performed the bolometer measurements. I.G., M.D. and S.B. characterized the temperature sensors and thin-film heaters. M.K., J.R., I.G., S.B.-G., S.B., T.Se. and T.So. analysed data and prepared figures with contributions from all authors. M.K., J.R., S.B., T.Se. and T.So. wrote the manuscript with comments from all authors. T.Se., S.B. and T.So. supervised the project and advised on device optimization.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.So. (someya@ee.tu-tokyo.ac.jp) or M.K. (martin@ntech.tu-tokyo.ac.jp).

METHODS

Ultrathin film substrate. PET sheets (thickness 125 μm) covered with a thin polydimethylsiloxane layer served as supporting substrates to ease handling of the 1.2- μm PEN foil for laboratory-scale device fabrication. The PEN film (obtained from Teijin Dupont Films, Japan) adheres to the supporting substrate only by weak van der Waals forces, which readily enables all subsequent processing steps (vacuum evaporation, anodization and chemical vapour deposition). After fabrication, the devices can be peeled off the supporting stack without causing damage.

Thin-film bolometers, temperature sensors and heating elements. Thin-film bolometers were fabricated by evaporating a resistive layer of aluminium 100 nm thick with 1 mm diameter through a shadow mask onto the 1.2- μm PEN substrate. The bolometers have a small heat capacity $C = 8 \mu\text{J K}^{-1}$, and the temperature coefficient of the resistance (TCR) of the aluminium electrode is $0.06\% \text{K}^{-1}$. Efficient heat transfer from the thin polymer foil to the copper block heat sink is facilitated by a thin layer of a liquid eutectic gallium–indium–tin alloy (galinstan), thus demonstrating efficient heat management strategies vital for high-power plastic electronic devices³¹. The bolometer was connected to a constant-current source delivering a current of 16 mA through the resistor. The resistance change was measured as a voltage decrease in a four-point probe technique configuration³², and the temporal voltage decrease was recorded directly with a Tektronix TDS 520C oscilloscope after amplification by a factor of 30 with an EG&G 5113 voltage amplifier. The bolometer was calibrated by measuring the TCR of the resistance element on a heating stage. Thermal pulse excitation was performed with a Nd:YAG laser at 1,064 nm with a pulse length of 10 ns and an adjustable pulse energy between 0 and 10 mJ. The transient signals of the bolometer after pulse excitation were fitted (according to equations (1) and (2) in Supplementary Information) with a thermal diffusivity of $D = (7.4 \pm 0.7) \times 10^{-8} \text{ m}^2 \text{ s}^{-1}$ (typical of PEN polymers). The temperature prefactor ΔT_0 was adjusted to obtain a best fit to the transients, owing to slight variations in the absorbed energy for the consecutive experiments on free-standing films and films coupled to the heat sink. The experiments show that heat is transferred to the heat sink very rapidly—on a time scale of only $4\tau = 9 \mu\text{s}$ (equation (2) in Supplementary Information). For the thin-film thermometers and resistive heating elements, gold strips 100 nm thick were thermally evaporated through a shadow mask at a pressure of 10^{-6} mbar and a rate of 1 nm s^{-1} . The strips were 1 mm wide and 1 cm long. Thin copper wires were attached to both ends of the gold strips for four-point resistivity measurements with a Keithley 2400 source meter. Thermal images were recorded by a NEC Thermotracer 7700 with 320×240 pixel resolution.

Thin-film transistors. The gate electrode was prepared by thermal evaporation of 40-nm aluminium through a shadow mask at a pressure of 10^{-6} mbar and a rate of 10 nm s^{-1} . The high- k aluminium oxide gate dielectric layer was formed by potentiostatic anodization under ambient conditions. A 0.01 M citric acid monohydrate ($\text{C}_6\text{H}_8\text{O}_7 \cdot \text{H}_2\text{O}$) electrolyte was prepared with ultrapure (18 M Ω cm) water. The contacted gates were then immersed in the electrolyte to form the working electrode (anode), and a platinum foil served as the counterelectrode. The thickness of the anodic oxide d_{ox} was determined by the applied anodization potential U according to $d_{\text{ox}} = k(U - U_{\text{ox}})$. The oxide formation factor k is 1.6 nm V^{-1} for aluminium^{17,18}. U_{ox} accounts for the naturally occurring passivation layer, which is about 2 nm, or $U_{\text{ox}} = -1.35 \text{ V}$ for aluminium. Anodic oxide formation on valve metals is a self-inhibiting process; thus, it is self-healing and converges to a uniform film thickness for a given potential. Optimal oxide quality is obtained when the anodization current drops to a steady-state value of about 1 μA for our geometry, and is typically reached within 10 min. Because the liquid electrolyte readily wets even a rough, bumpy surface, electrochemical oxidation is a highly reliable method of forming a dense dielectric layer, which ultimately decreases leakage currents and enables high device yields. To demonstrate the precise scalability of the dielectric thickness and thus the operating voltage of the transistor, devices were fabricated with oxide 9 nm and 19 nm thick (operated at up to 3 V and up to 5 V, respectively). We functionalized the aluminium oxide surface with an n-tetradecyl phosphonic acid SAM³³ to form a suitable interface to the organic semiconductor DNTT and decrease the interface trap density. After brief surface activation in oxygen plasma (1 min at 100 W), the gates were dipped into a 5 mM solution of the SAM molecules in propan-2-ol to form a SAM 2 nm thick. Our hybrid gate dielectric layer has interface capacitances of about 420 nF cm^{-2} and about 280 nF cm^{-2} for devices with 9 nm alumina and 19 nm alumina, respectively. Subsequently, 20 nm DNTT was evaporated through a shadow mask at a pressure of 10^{-6} mbar and a rate of 0.1 Å s^{-1} . Gold 30 nm thick evaporated through a shadow mask at a pressure of 10^{-6} mbar and a rate of 1 Å s^{-1} formed source and drain electrodes with width 500 μm and length 40 μm for single transistors. We note that our transistor fabrication process is compatible with n-type or ambipolar organic semiconductors^{34–37}, which may further expand possible applications of imperceptible electronics by taking full advantage of organic complementary circuits and device architectures^{6,38–40}.

Active-matrix tactile sensing foil. The fabrication methods employed for a single transistor are readily scalable and require no additional process steps for the 12×12 active-matrix transistor backplane. Transistors with a channel width of 12,000 μm and length of 20 μm provide high on-currents at low operating voltage. For the active-matrix transistors, 19-nm anodic alumina was formed. We deposited parylene diX-SR (about 800 nm thick; Daisan Kasei Co. Ltd) by chemical vapour deposition to separate the switching matrix backplane from the sensor layer. A green laser marker system (T-Centric SHG Laser Marker; Keyence) was used to drill precise, high-resolution vias to the underlying transistors (the typical via diameter was 30 μm). Interdigitated gold electrodes 100 nm thick with a 30 μm pitch were evaporated through a shadow mask at a pressure of 10^{-6} mbar and a rate of 1 Å s^{-1} , and served as resistive tactile sensors. To perceive pressure signals from non-conducting objects, a pressure-sensitive rubber sheet (CS57-7RSC; PCR Technical Corp.) was laminated on top of the tactile foil.

Neutral mechanical plane calculation. The position b of the neutral mechanical plane (zero-strain position) of a multilayer stack with the n th layer on top of the first layer on the bottom is given by⁸

$$b = \frac{\sum_{i=1}^n E_i t_i \left[\sum_{j=1}^i t_j - \frac{t_i}{2} \right]}{\sum_{i=1}^n E_i t_i} \quad (1)$$

where E_i and t_i denote the Young's moduli and thicknesses of the individual layers. For the ultrathin OFETs ($n = 6$; 1.2 μm PEN/25 nm Al/19 nm anodic AlO_x /20 nm DNTT/40 nm Au/0.8 to 1 μm parylene diX-SR) the elastic moduli are $E_{\text{PEN}} = 4 \text{ GPa}$, $E_{\text{Al}} = 69 \text{ GPa}$, $E_{\text{AlO}_x} = 130 \text{ GPa}$, $E_{\text{DNTT}} = 2.3 \text{ GPa}$, $E_{\text{Au}} = 79 \text{ GPa}$ and $E_{\text{diX-SR}} = 4 \text{ GPa}$. This places the neutral mechanical plane at a distance b of 1.15–1.2 μm from the bottom surface of the PEN substrate, near the position of the active OFET structure.

Compression and stretching. Transistors for the compression/stretching experiments were fabricated as described above for single transistors but with extended gate, source and drain contacts to enable electrical connection for the bending and stretching tests. The devices were encapsulated with a parylene layer 0.8–1 μm thick that placed the transistors near the neutral plane and protected them from mechanical damage. Transistors for the stretching/compression tests had a channel width of 1,000 μm and length of 100 μm . A schematic diagram of the compression/stretching setup together with a picture of the setup is shown in Supplementary Fig. 17. A strip of 3M VHB 4905 elastomer was fixed between two clamps and pre-stretched. Polyimide tape was used as a rigid delimiter to define the compression and stretching zone, and the distance between the two rigid delimiters was measured to be exactly 1 cm (defined as L_0). The transistors were then placed on the elastomer. The active device area (gate area) of the transistor was $1.75 \times 1.5 \text{ mm}^2$. Together with the gate, source and drain contacts, the transistor occupied roughly 65% of the 1 cm gap in the compression direction. The transistor was then compressed by moving the rigid delimiters closer together, thereby allowing the elastomer to relax, and compressing the device area. The compression zone measured $2 \times 1 \text{ cm}^2$, meaning that the deformation was quasi-linear and quasi-reversible, although there was always a small, unintentional biaxial and/or permanent deformation due to the viscoelasticity of the elastomer. We defined the compression as $(L_0 - L)/L_0$. Shrinking the gap from 10 mm to 3 mm corresponded to a 70% compression. The definition of tensile strain used the fully relaxed (smallest) state as the starting point. After relaxation to 3 mm, re-stretching to 10 mm gave a tensile strain of $(10 - 3 \text{ mm})/3 \text{ mm}$; that is, 233% tensile strain. For the biaxially stretchable transistors, the elastomer was fixed at all sides and pre-stained by pressing a 1-cm cylinder against the bottom surface. The transistor foil was then applied to the top surface and the cylinder was lowered, relaxing the elastomer. The change in device area was measured by monitoring circular markers on the elastomer.

Device characterization. Electrical characterization of transistors and active-matrix sensing sheets was performed with a Semiconductor Parameter Analyser (Agilent B1500A) under ambient laboratory conditions. Thermal stability tests were performed under ambient atmospheric conditions. Immersion tests in aqueous solution were performed in saline solution (0.9wt% NaCl; Becton Dickinson and Company prepared saline solution, normal (physiological)). Optical micrographs and three-dimensional topological images were recorded with a Keyence VK-9710 ultraviolet laser microscope and a Keyence VHX-2000 digital microscope. SEM images of wrinkled transistors on the elastic substrate were taken with a Hitachi S-4800 Field Emission Scanning Electron Microscope at an acceleration voltage of 20 kV after fixing the samples onto a rigid glass substrate and depositing a gold layer less than 5 nm thick. Cross-sectional SEM (S-5200 Ultra-High Resolution FE SEM) and high-resolution STEM (200-kV HD-2700 Cs-corrected STEM) images were taken by Hitachi High-Technologies Corporation. Surface morphologies were characterized with a NanoScope IIIa (Veeco Instruments) atomic force microscope.

31. Gärditz, C., Winnacker, A., Schindler, F. & Paetzold, R. Impact of Joule heating on the brightness homogeneity of organic light emitting devices. *Appl. Phys. Lett.* **90**, 103506 (2007).
32. Bauer, S. & Ploss, B. A method for the measurement of the thermal, dielectric, and pyroelectric properties of thin pyroelectric films and their applications for integrated heat sensors. *J. Appl. Phys.* **68**, 6361–6367 (1990).
33. Facchetti, A., Yoon, M. H. & Marks, T. J. Gate dielectrics for organic field-effect transistors: new opportunities for organic electronics. *Adv. Mater.* **17**, 1705–1725 (2005).
34. Yan, H. *et al.* A high-mobility electron-transporting polymer for printed transistors. *Nature* **457**, 679–686 (2008).
35. Usta, H., Facchetti, A. & Marks, T. J. Air-stable, solution-processable n-channel and ambipolar semiconductors for thin-film transistors based on the indenofluorenebis(dicyanovinylene) core. *J. Am. Chem. Soc.* **130**, 8580–8581 (2008).
36. Oh, J. H. *et al.* High-performance air-stable n-type organic transistors based on core-chlorinated naphthalene tetracarboxylic diimides. *Adv. Funct. Mater.* **20**, 2148–2156 (2010).
37. Jung, B. J., Lee, K., Sun, J., Andreou, A. G. & Katz, H. E. Air-operable, high-mobility organic transistors with semifluorinated side chains and unsubstituted naphthalenetetracarboxylic diimide cores: high mobility and environmental and bias stress stability from the perfluorooctylpropyl side chain. *Adv. Funct. Mater.* **20**, 2930–2944 (2010).
38. Zhang, X.-H., Potscavage, W. J., Choi, S. & Kippelen, B. Low-voltage flexible organic complementary inverters with high noise margin and high dc gain. *Appl. Phys. Lett.* **94**, 043312 (2009).
39. Geib, S. *et al.* Core-brominated tetraazaperopyrenes as n-channel semiconductors for organic complementary circuits on flexible substrates. *Adv. Funct. Mater.* <http://dx.doi.org/10.1002/adfm.201203600> (27 May 2013).
40. Baeg, K.-J. *et al.* High speeds complementary integrated circuits fabricated with all-printed polymeric semiconductors. *J. Polym. Sci. B* **49**, 62–67 (2010).

Heavy solitons in a fermionic superfluid

Tarik Yefsah¹, Ariel T. Sommer¹, Mark J. H. Ku¹, Lawrence W. Cheuk¹, Wenjie Ji¹, Waseem S. Bakr¹ & Martin W. Zwierlein¹

Solitons—solitary waves that maintain their shape as they propagate—occur as water waves in narrow canals, as light pulses in optical fibres and as quantum mechanical matter waves in superfluids and superconductors. Their highly nonlinear and localized nature makes them very sensitive probes of the medium in which they propagate. Here we create long-lived solitons in a strongly interacting superfluid of fermionic atoms and directly observe their motion. As the interactions are tuned from the regime of Bose–Einstein condensation of tightly bound molecules towards the Bardeen–Cooper–Schrieffer limit of long-range Cooper pairs, the solitons’ effective mass increases markedly, to more than 200 times their bare mass, signalling strong quantum fluctuations. This mass enhancement is more than 50 times larger than the theoretically predicted value. Our work provides a benchmark for theories of non-equilibrium dynamics of strongly interacting fermions.

Solitonic excitations are found throughout nature, in proteins and DNA, in crystals as dislocations, and generally in the wake of symmetry-breaking phase transitions¹. In fermionic systems, solitonic defects may provide bound states for fermions that often play a crucial role in the system’s transport properties. Famous examples are Andreev bound states inside vortex cores², fractionally charged solitons in relativistic quantum field theory^{3,4}, and the spinless charged solitons responsible for the high conductivity of polymers⁵. However, the free motion of such defects in electronic systems is hindered by pinning at impurities^{5,6}. Fermionic superfluids of ultracold atomic gases provide a paradigmatic form of quantum matter^{7–9} that is free of impurities, where—as we show here—real-time dynamics of solitonic excitations can be directly observed.

Superfluids are described by a complex macroscopic wavefunction that is rigid against twists of its phase. The ground state of the superfluid thus has uniform phase, and small perturbations propagate as sound waves. A nonlinear excitation—the dark soliton—occurs when the phase is twisted substantially over a short range. In the extreme case of a phase jump by 180°, the wavefunction changes sign and crosses zero at the location of the jump, creating a stationary black soliton. In weakly interacting Bose–Einstein condensates (BECs) all bosons reside in the condensate, so the particle density vanishes at a black soliton, and is reduced for a moving dark soliton. Solitons in BECs have been studied extensively both theoretically and experimentally¹⁰. In a series of pioneering experiments, dark solitons have been created via phase-imprinting^{11–14} or in the wake of shock waves^{15–17}. Collisions of two dark solitons¹⁷ and soliton oscillations^{14,17} were observed. Solitons in weakly interacting BECs are well described as solutions to the Gross–Pitaevskii equation, known in other contexts as the cubic nonlinear Schrödinger equation.

In fermionic superfluids^{7–9}, solitons are phase twists in the wavefunction of fermion pairs^{18,19}. For *s*-wave superfluids, the pair wavefunction is also known as the pairing gap $\Delta(r)$, which in general can depend on the spatial location *r*. By tuning the interactions between fermions, one can access the crossover from Bose–Einstein condensation of molecules to the Bardeen–Cooper–Schrieffer (BCS) state of long-range Cooper pairs. In the limit of tight molecular pairing, interactions between molecules are weak and the molecular condensate is still described by the Gross–Pitaevskii equation. Stationary solitons are thus again devoid of particles. In this limit, the wavefunction for a stationary soliton, shown in Fig. 1a, depends on position along the *z*

axis as $\Delta(z) = \Delta_0 \tanh(z/\xi)$, where Δ_0 is the magnitude of the wavefunction in the bulk, far away from the soliton, and the soliton width ξ is equal to the healing length of the condensate²⁰. The repulsive interactions between the molecular bosons can be increased by means of a Feshbach resonance, allowing the study of strongly interacting Bose gases⁷. Strong interactions increase the importance of quantum fluctuations that are present even at zero temperature, leading to a depletion of the condensate. The uncondensed bosons are expected to fill in the soliton notch, the void at the soliton’s position, in order to minimize their repulsive interaction with the condensate^{10,21–24}. Figure 1a shows the density profile of the bosons localized at the soliton, the so-called anomalous mode that is predicted to be the main contribution to the density inside the soliton notch^{21–23}. Similar soliton filling has been predicted for BECs in optical lattices, where the effect of interactions and thus the role of quantum fluctuations is enhanced by reducing the particles’ kinetic energy²⁵.

Description of solitons in the BEC–BCS crossover

When the interaction strength in the pair condensate becomes of the order of the Fermi energy (E_F), the composite nature of the molecules is revealed. The fermion pair size is now of the order of the interparticle spacing, and the system is a crossover superfluid in between the BEC and BCS limits of superfluidity^{7–9}. A unified description for solitons in fermionic superfluids throughout the BEC–BCS crossover has been found within mean-field theory via the Bogoliubov–de Gennes (BdG) equation for a spatially varying gap $\Delta(z)$ (refs 19, 26–28):

$$\left\{ \left(-\frac{\hbar^2 \nabla^2}{2m} - \mu \right) \sigma_z + \Delta(z) \sigma_x \right\} \begin{pmatrix} u_n \\ v_n \end{pmatrix} = E_n \begin{pmatrix} u_n \\ v_n \end{pmatrix} \quad (1)$$

where \hbar is Planck’s constant *h* divided by 2π , *m* is the atomic mass, μ is the chemical potential, $\sigma_{x,y,z}$ are Pauli matrices, and $u_n(z)$ and $v_n(z)$ are the amplitudes describing the particle and hole character of Bogoliubov quasi-particles of energy E_n (we omit spin indices). The order parameter $\Delta(z)$ is related to the quasi-particle amplitudes by the self-consistency relation $\Delta(z) = -g \sum_n u_n(z) v_n^*(z)$, where *g* is the coupling strength, tunable via the scattering length *a* between fermions, and * denotes complex conjugation. The BdG equations have been shown to reduce to the Gross–Pitaevskii equation for bosonic molecules in the BEC limit²⁹, where stationary solitons are devoid of particles. As the interactions are tuned from the BEC to the BCS regime, the BdG equations predict an increasing filling of the soliton¹⁹. At the

¹MIT–Harvard Center for Ultracold Atoms, Research Laboratory of Electronics, Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

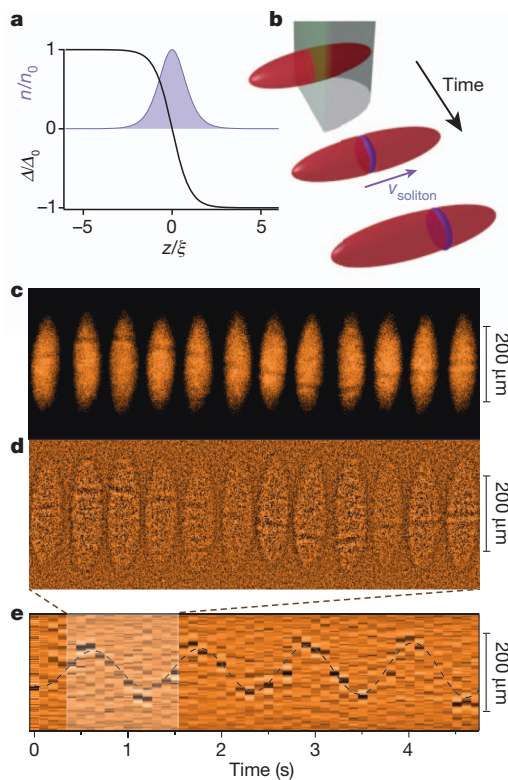


Figure 1 | Creation and observation of solitons in a fermionic superfluid.

a, Superfluid pairing gap $\Delta(z)$ for a stationary soliton, normalized by the bulk pairing gap Δ_0 , and density $n(z)$ of the localized bosonic (fermionic) state versus position z , in the BEC (BCS) regime of the crossover, in units of the BEC healing length (BCS coherence length) ξ . **b**, Diagram of the experiment. A phase-imprinting laser beam twists the phase of one-half of the trapped superfluid by approximately π . The soliton generally moves at non-zero velocity v_{soliton} . **c**, Optical density and **d**, residuals (optical density minus a smoothed copy of the same image) of atom clouds at 815 G, imaged via the rapid ramp method³⁴, showing solitons at various hold times after creation. One period of soliton oscillation is shown. The in-trap aspect ratio was $\lambda = 6.5(1)$. **e**, Radially integrated residuals as a function of time revealing long-lived soliton oscillations. The soliton period is $T_s = 12(2)T_z$, much longer than the trapping period of $T_z = 93.76(5)$ ms, revealing an extreme enhancement of the soliton's relative effective mass, M^*/M .

Feshbach resonance, in the unitarity limit where the scattering length diverges, a substantial part of this filling is due to so-called Andreev bound states, localized fermionic states bound to the soliton, also known to reside inside vortex cores³. Here, the gas density in the vicinity of the soliton is predicted to be suppressed by 80% of the bulk density, as opposed to 100% for solitons in BECs.

In the BCS limit of weak attractive interactions, the BdG equations reduce to the Andreev equation, a Dirac equation where the pairing gap $\Delta(z)$ plays the role of a spatially varying mass coupling particles and holes³ (see Supplementary Information). The same equation describes solitons in conducting polymers⁵. The solution for the pairing gap is known⁵ to be $\Delta(z) = \Delta_0 \tanh(z/\xi_{\text{BCS}})$, as in the BEC limit, that is, it is again represented by Fig. 1a but with $\xi = \xi_{\text{BCS}}$, the BCS coherence length. The density profile of the localized state in Fig. 1a here represents the fermionic Andreev bound state, as opposed to the density of uncondensed bosons in the BEC regime. Solitons in the BCS regime are expected to be essentially completely filled in. Indeed, in this limit of long-range overlapping Cooper pairs, only a minute fraction of particles near the Fermi surface takes part in pairing, and the reduction of the pairing gap at the soliton affects the density only very weakly.

Creating solitons in a fermionic superfluid

The creation of solitons in a strongly interacting fermionic superfluid poses several challenges. First, a superfluid with a soliton is not in its

ground state, so the temperature of the gas has to be low enough for the soliton not to decay rapidly into thermal excitations. Such dissipation can proceed through collisions of the soliton with sound waves, leading to its acceleration. When the soliton reaches a critical velocity, it is expected to decay into phonons or, in the case of fermionic superfluids, pair excitations^{27,28,30}. Second, solitons can generally decay into vortices via the so-called snake instability^{13,15,31,32}. In the case of weakly interacting BECs in elongated traps, stability requires the chemical potential μ of the condensate to be not much larger than the transverse confinement energy³¹. For a Fermi gas, this would require a quasi-one-dimensional geometry where the transverse cloud width is one interparticle spacing. As we show below, this is not necessary. Last, for strongly interacting superfluids, it is a priori not obvious that solitons are stable against quantum fluctuations^{10,21–25,33}.

Here we create and observe long-lived solitons in a strongly interacting fermionic superfluid of ^6Li atoms near a Feshbach resonance. Solitons are created via phase imprinting (see Fig. 1b), a technique successfully employed for weakly interacting Bose condensates^{11,12,14}. The superfluid containing typically $\sim 2 \times 10^5$ atom pairs is prepared in an elongated trap with cylindrical symmetry (axial and radial trapping period respectively $T_z = 45\text{--}210$ ms and $T_\perp = 14$ ms) and tunable aspect ratio $\lambda = T_z/T_\perp$ (ref. 7). A green laser beam far detuned from the atomic resonance is masked to shine on one half of the superfluid. In a time t , the applied potential U , as experienced by a single fermion, advances the phase of the superfluid order parameter in the exposed region by $\Delta\phi = 2Ut/\hbar$ relative to the unexposed region. The time $t \approx 35 \mu\text{s}$ is experimentally adjusted in order to create one high-contrast soliton.

In the strongly interacting regime, the soliton does not cause a density depletion within our resolution. However, it is tied to a phase twist in the pair wavefunction. As in the case of vortices³⁴, the pair wavefunction can be directly observed via a rapid ramp to the BEC side of the Feshbach resonance. The ramp converts large fermion pairs into tightly bound molecules, empties out the soliton cores and increases the soliton width to the final healing length $\propto 1/\sqrt{n_M a_f}$, where a_f is the scattering length at the final magnetic field and n_M the density of molecules. The rapid ramp followed by time-of-flight expansion thus enhances the soliton contrast and acts as a magnifying glass (for details, see Supplementary Information).

Figure 1c and d report the observation of solitons in a fermionic superfluid prepared at 815 G (close to the 832 G Feshbach resonance) for various hold times following the phase imprint. Here, the interaction parameter at the cloud centre is $1/k_F a = 0.30(2)$, where a is the scattering length and $k_F = (3\pi^2 n)^{1/3}$ is the Fermi wavevector, related to the total central fermion density n and the Fermi energy $E_F = \hbar^2 k_F^2/2m$. Figure 1c shows the optical density in absorption images taken after time of flight and the rapid ramp to ~ 580 G, while Fig. 1d displays residuals obtained by subtracting a smoothed copy of the same absorption image. The optical density contrast of solitons is about 10% (see Supplementary Information). A sequence of radially integrated residuals as a function of time is displayed in Fig. 1e, demonstrating the soliton to be stable for more than 4 s or 100,000 times the microscopic timescale \hbar/E_F , the Fermi time. This establishes that solitons in fermionic superfluids can exist as stable and long-lived excitations that do not decay despite strong quantum fluctuations.

Soliton oscillations

The solitons are observed to undergo oscillations in the harmonically trapped superfluid, demonstrating their emergent particle nature. The motion is to a high degree deterministic, as soliton positions for different realizations of the experiment at varying wait times lie on the same classical sinusoidal trajectory. The force on the soliton is provided by the trapping force experienced by the atoms missing in the soliton, $N_s m \omega_z^2 z \equiv M \omega_z^2 z$, where $\omega_z = 2\pi/T_z$, $|N_s|$ is the number of missing atoms, and $M = N_s m < 0$ the bare mass of the soliton. M is negative as the soliton is a density depletion. Introducing the effective, or

inertial mass of the soliton M^* , this force causes an acceleration $\ddot{z} = -\frac{M}{M^*}\omega_z^2 z$. Because we observe oscillations, M^* must be negative as well, implying that the soliton is an effective particle that decreases its kinetic energy as it speeds up. One obtains a direct relation²⁶ between the relative effective mass M^*/M and the normalized soliton period T_s/T_z :

$$\frac{M^*}{M} = \left(\frac{T_s}{T_z}\right)^2 \quad (2)$$

The observed soliton period of oscillation T_s is about one order of magnitude longer than the trapping period T_z for single atoms. This directly indicates an extreme enhancement of the relative effective mass. In general, the difference between the effective mass M^* and the bare mass M of the soliton arises from the phase slip $\Delta\phi$ across the soliton, which implies a superfluid counterflow²⁶. For the soliton to move, an entire sheet of atoms thus has to flow past it. The difference $M - M^*$ is the mass of that sheet, given by the mass density multiplied by the entire soliton volume. In contrast, the soliton's bare mass M is only due to the mass deficit of $|N_s|$ atoms and can become much smaller in magnitude than M^* when the soliton is filled. For weakly interacting BECs, where solitons are devoid of particles, the effective mass is still of the same order of the bare mass, $(M^*/M)_{\text{BEC}} = 2$. This leads to an oscillation period that is only $\sqrt{2}$ times longer than T_z (refs 20, 35), as has been observed in experiments^{14,17}. In the BCS limit, where only a minute fraction Δ_0/E_F of the gas contributes to Cooper pairing, $|N_s| \propto \Delta_0/E_F \propto \exp[-\pi/(2k_F|a|)]$ and thus the soliton's relative effective mass can be expected to become exponentially large.

Indeed, as shown in Fig. 2, we find that the soliton period, and hence the relative effective mass, increases dramatically as the interactions are tuned from the limit of Bose–Einstein condensation (Fig. 2a) towards the BCS limit. At 700 G, where $1/k_F a = 2.6(2)$, the system represents a strongly interacting Bose gas of molecules⁷. The soliton period is $T_s = 4.4(5)T_z$, already three times longer than in the case of a weakly interacting BEC. At the Feshbach resonance (Fig. 2d), we measure a soliton period of $T_s = 14(2)T_z$, corresponding to a relative effective mass of $M^*/M = 200(50)$. This is more than 50 times larger than the

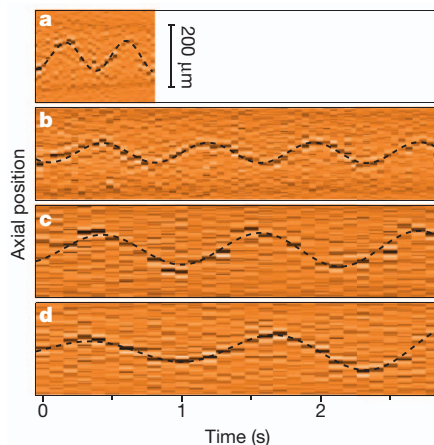


Figure 2 | Soliton oscillations in the BEC–BCS crossover. Shown are soliton oscillations in a trapped fermionic superfluid for various magnetic fields B around the Feshbach resonance. **a–d**, The soliton period is observed to markedly increase as the system is tuned from the BEC regime (**a**) to the Feshbach resonance (**d**). The measured period (T_s/T_z), magnetic field (B in G) and interaction parameter at the cloud centre $1/k_F a$ were respectively: **a**, 4.4(5), 700, 2.6(2); **b**, 7.5(9), 760, 1.4(1); **c**, 12(2), 815, 0.30(2); **d**, 14(2), 832, 0. The initial atom number per spin state (N_0), its decay rate (τ in s) and Thomas–Fermi radius after time of flight (R_{TF} in μm) range respectively from: 1.1×10^5 , 1.2(2), 135 at $B = 700$ G to 2.3×10^5 , 12(1) and 200 on resonance. The aspect ratio is $\lambda = 6.2(7)$. Note that at $B = 700$ G, the superfluid is short lived due to enhanced three-body loss. At 760 G (**b**), the soliton survived for more than 6 s, comparable to the lifetime of the superfluid at that field.

result of mean-field BdG theory in three dimensions^{26,36} that predicts $M^*/M = 3$. Note that the superfluid is fully three-dimensional: on resonance, the chemical potential $\mu \approx 35\hbar\omega_\perp$, where ω_\perp is the radial trapping frequency. Still, for very elongated traps, one expects to reach a universal quasi-one-dimensional regime where the tight radial confinement is irrelevant for propagation along the long axis³⁷. This prompted us to study the dependence of the soliton period on the aspect ratio of our trap.

Figure 3 summarizes our measurements for the soliton period and the relative effective mass as a function of the interaction parameter $1/k_F a$ throughout the BEC–BCS crossover, for aspect ratios $\lambda = 3.3$, 6.2 and 15. The strong increase of M^*/M towards the BCS regime is observed for all trap geometries. The normalized soliton period T_s/T_z appears to converge to a limiting value for the most elongated trap: the normalized period changes by only 15% as the aspect ratio is increased by more than a factor of two from 6.2 to 15. This indicates that the soliton dynamics approach a universal quasi-one-dimensional limit. Even in a much less elongated trap with $\lambda = 3.3(1)$, the soliton period is only slightly increased by about 30% compared to $\lambda = 6.2$, accompanied by an increased susceptibility of the soliton towards bending or ‘snaking’^{10,13,15} (for examples, see Supplementary Information).

We attribute the large relative effective mass M^*/M in the strongly interacting regime to the filling of the soliton with uncondensed fermion pairs resulting from strong quantum fluctuations. Similar filling with uncondensed particles has been predicted for solitons in strongly interacting Bose condensates^{10,22–25,33}. A substantial filling of the soliton will reduce the number $|N_s|$ of atoms missing inside the soliton, therefore considerably weaken the restoring harmonic force from the trap and strongly increase M^*/M . At the Feshbach resonance, our *in situ* density profiles provide a lower bound on the soliton filling of 90%, compared to the expected 20% from mean-field theory (see Supplementary Information). Mean-field theory for the BEC–BCS crossover heavily underestimates the role of quantum fluctuations already on the BEC side, where it predicts a fraction of uncondensed bosons that scales as na^3 instead of the correct $\sqrt{na^3}$ scaling⁷. Our experiment thus directly reveals the importance of beyond mean-field effects for the dynamics of strongly interacting fermionic superfluids. Significant soliton filling was found theoretically in a strongly interacting relativistic superfluid using methods from string theory^{38–40}. For the resonantly interacting Fermi gas, a theoretical study based on a

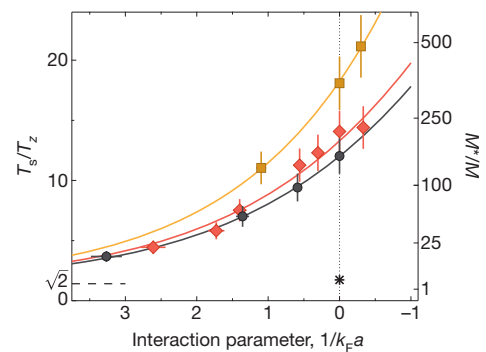


Figure 3 | Soliton period and effective mass versus interaction strength in the BEC–BCS crossover. The normalized soliton period T_s/T_z is shown as a function of the interaction parameter $1/k_F a$ in the cloud centre, for three different trap aspect ratios: $\lambda = 15(1)$ (black circles), $6.2(7)$ (red diamonds) and $3.3(1)$ (orange squares). The error bars correspond to the typical spread over five measurements, and the solid lines are guides to the eye. The soliton period strongly increases from the BEC regime towards the Feshbach resonance (vertical dotted line), where $T_s/T_z = 12(2)$ for $\lambda = 15(1)$, and to the BCS side. This directly reflects an extreme enhancement of the relative effective mass $M^*/M = T_s^2/T_z^2$, which we attribute to strong quantum fluctuations and filling of Andreev bound states. The result for a weakly interacting BEC, $T_s/T_z = \sqrt{2}$, is shown as the dashed line. The star marks the mean-field prediction²⁶ at unitarity $M^*/M = T_s^2/T_z^2 = 3$.

density functional approach found solitons with clear filling in the wake of shock waves⁴¹. The strong increase of the soliton period is reminiscent of the situation for dark-bright solitons in weakly interacting BECs, where a distinguishable atomic species or another spin state resides inside the soliton notch^{14,42,43}. For fermions, mean-field theory in the strongly interacting regime attributes a substantial part of the soliton filling to Andreev bound states^{19,26–28}. These are also predicted to carry the dominant fraction of the superfluid flow across the soliton, which can be regarded, in its rest frame, as a Josephson junction of vanishing barrier height²⁹. It will be an interesting topic for future experiments to determine the contribution of Andreev states to the soliton filling.

Temperature dependence

To demonstrate that the slow soliton oscillations are a truly quantum effect and not due to the finite temperature of our gas, we investigated the soliton motion as a function of temperature for the unitary Fermi gas at the Feshbach resonance (Figs 4 and 5). A measure of temperature is provided by the thermal fraction, the number of uncondensed molecules observed after the rapid ramp. The soliton period is found to be insensitive to changes in temperature within the measurement uncertainty (Fig. 5a).

The stability of solitons is, however, strongly affected by thermal effects. At low temperatures, the soliton oscillation occurs essentially without energy loss, demonstrating dissipationless flow (Fig. 4a). For increasing temperature, we observe anti-damping of soliton oscillations (Fig. 4b). This is characteristic of a particle with negative mass that can lower its energy by accelerating. To our knowledge, such anti-damping of solitons has not been directly observed previously in a quantum gas experiment. The energy loss is likely to be due to collisions with thermally induced phonons¹⁰, and we indeed observe a strong decrease in the anti-damping time constant as the temperature is raised (Fig. 5b). At even higher temperatures, the soliton's position becomes less reproducible (Fig. 4c) and its lifetime is strongly reduced (Fig. 5c). Concurrently, we observe increased axial fluctuations in the superfluid (see Fig. 4d–f), some of which appear to have comparable contrast to the imprinted soliton. These additional solitons might be

'thermal solitons', predicted to occur even in equilibrium in weakly interacting Bose condensates⁴⁴. Similar to vortex–anti-vortex pairs in two dimensions, soliton–anti-soliton pairs can be expected to spontaneously break in one dimension and proliferate.

We note that on resonance, the fastest solitons we observe move at the exceedingly slow speed of 0.50 mm s^{-1} or 5% of the (independently measured) speed of sound on resonance. Their sudden disappearance, observed for example in Fig. 4c, can thus not be related to motion close to the Landau critical speed. Instead, their decay might be tied to inelastic collisions with thermal solitons, as soliton collisions have been found to become increasingly inelastic towards the BCS side in theoretical simulations²⁸. Another possibility for their decay at such low speeds is that the soliton's energy dispersion has a minimum at an unexpectedly small fraction of the critical velocity²⁸. One might expect fermion pairs to break at finite temperatures and fill in the soliton, in addition to quantum fluctuations. However, even for the highest thermal fraction where solitons have been observed, the actual temperature is determined to be below $T = 0.10 E_F / k_B$ (k_B is the Boltzmann constant), while the bulk pairing gap is about $\Delta_0 = 0.4 E_F$ (ref. 45). Pair breaking should thus still be exponentially suppressed, explaining the insensitivity of the soliton period to the thermal fraction.

Conclusion and outlook

We have created and observed long-lived solitons in a strongly interacting fermionic superfluid. Their period of oscillation and thus their relative effective mass increases markedly as the interactions are tuned from the BEC limit of tightly bound molecules towards the BCS limit of long-range Cooper pairs. This signals strong, beyond mean-field, effects, which are likely to be due to uncondensed fermion pairs filling the soliton, in addition to purely fermionic Andreev bound states. Our study provides an important quantitative benchmark for theories of non-equilibrium dynamics of strongly interacting Fermi gases. An exciting prospect is to directly detect the Andreev bound states spectroscopically^{19,46}. Although they are not topologically protected, their lifetime should equal that of the soliton—many seconds or 100,000

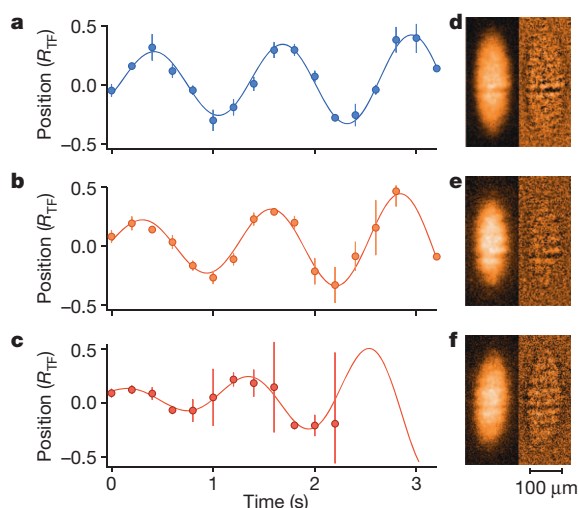


Figure 4 | Soliton motion in the unitary Fermi gas at various temperatures. a–c, Soliton trajectories for increasing temperature, with thermal fractions a, 7(2)%, b, 9(2)% and c, 15(3)%. The error bars indicate the standard deviation of typically five repetitions and the solid lines are fits to the data to the anti-damped sinusoidal function $f(t) \propto \exp(t/\tau_s) \sin(2\pi t/T_s + \phi)$. Whereas the period is found to be independent of temperature within our uncertainty, the anti-damping time decreases from $\tau_s/T_s = 5(2)$ for the coldest clouds (a) to $\tau_s/T_s = 1.3(5)$ for the hottest ones (c). d–f, Representative optical densities (left) and residuals (right) of the superfluid after the rapid ramp. Whereas at low temperatures, the soliton is the only significant density variation, at higher temperatures transverse stripes appear that we tentatively interpret as thermal solitons.

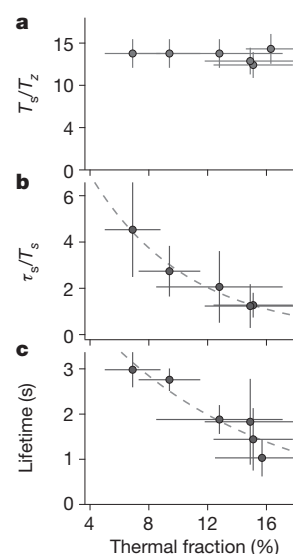


Figure 5 | Effect of finite temperature on soliton motion. a, The soliton period is found to be insensitive to temperature. b, The $1/e$ anti-damping time and c, the soliton lifetime, are found to be strongly dependent on the thermal fraction. The soliton lifetime is defined as the time when the probability of observing a soliton decreased to 50%. The dashed lines are guides to the eye. The horizontal error bars indicate the standard deviation of the thermal fraction within a data set. The vertical error bars in a represent the typical spread over five measurements, those in b result from the contribution of the fitting error on τ_s and the error on T_s , and those in c reflect the time difference between having 90% and 10% survival probability.

Fermi times—so that they might become a useful quantum resource. In the presence of spin imbalance, the soliton represents a limiting case of the long-sought Fulde–Ferrell–Larkin–Ovchinnikov state of moving Cooper pairs^{46–49}. Indeed, it is energetically favourable for an excess fermion to reside inside a soliton rather than inside the bulk superfluid. Although it is difficult to realize the Fulde–Ferrell–Larkin–Ovchinnikov state in equilibrium, direct engineering of soliton trains might produce a long-lived metastable analogue.

METHODS SUMMARY

Preparation. The atomic gas is composed of a balanced mixture of the two lowest hyperfine states of ⁶Li initially prepared at 760 G (ref. 7). The trapping potential results from the combination of a magnetic field curvature providing harmonic confinement in the axial direction and an optical dipole trap (wavelength 1,064 nm) providing tighter radial confinement. The axial periods are $T_z = 210$ ms, 95 ms and 45 ms, respectively, for the three aspect ratios considered here.

Phase imprinting. A step-like intensity profile is imprinted on a laser beam (wavelength 532 nm, power ~200 mW) by means of an opaque mask. A 3- μ m-resolution imaging system projects the intensity distribution at the mask location onto the atoms (the beam waist at the atoms is 60 μ m). A phase twist of π corresponds to a pulse time of about 30 μ s, much shorter than the timescale \hbar/μ associated with a typical chemical potential, of the order of few 100 μ s. The pulse duration is finely adjusted to yield exactly one soliton with high contrast observed after the rapid ramp. Because we imprint only a phase step but not a density depletion, sound waves must be generated in addition to the soliton^{11,12}. The sound waves are found to die out in a quarter axial trapping period when they have reached the edge of the atom cloud (see Supplementary Information). For the data above 760 G, the soliton is created at 760 G and the magnetic field is subsequently ramped (in about $2T_z$) to the final magnetic field where the soliton motion is studied. For final magnetic fields below 760 G, the soliton is created at that field. We found that solitons can be created directly at the Feshbach resonance as well.

Thermometry. Thermometry from fits of density profiles at the Feshbach resonance to the known equation of state⁵⁰ yielded an upper limit of $T = 0.05E_F/k_B = 10$ nK $\approx 3\hbar\omega_{\perp}$ for our lowest temperatures, where $\omega_{\perp} = 2\pi/T_{\perp}$. At such low temperatures, far below the critical temperature for superfluidity, thermometry via fitting is less sensitive to small changes in temperature⁷. We therefore use the thermal fraction of molecules observed after a rapid ramp to the BEC side of the Feshbach resonance as a robust thermometer⁷.

Received 19 February; accepted 28 May 2013.

Published online 17 July 2013.

- Zurek, W. Cosmological experiments in condensed matter systems. *Phys. Rep.* **276**, 177–221 (1996).
- Caroli, C., Gennes, P. d. & Matricorn, J. Bound fermion states on a vortex line in a type II superconductor. *Phys. Lett.* **9**, 307–309 (1964).
- Jackiw, R. & Rebbi, C. Solitons with fermion number 1/2. *Phys. Rev. D* **13**, 3398–3409 (1976).
- Goldstone, J. & Wilczek, F. Fractional quantum numbers on solitons. *Phys. Rev. Lett.* **47**, 986–989 (1981).
- Heeger, A. J., Kivelson, S., Schrieffer, J. R. & Su, W. P. Solitons in conducting polymers. *Rev. Mod. Phys.* **60**, 781–850 (1988).
- Monceau, P. Electronic crystals: an experimental overview. *Adv. Phys.* **61**, 325–581 (2012).
- Ketterle, W. & Zwerlein, M. Making, probing and understanding ultracold Fermi gases. *Riv. Nuovo Cim.* **31**, 247–422 (2008).
- Giorgini, S., Pitaevskii, L. P. & Stringari, S. Theory of ultracold atomic Fermi gases. *Rev. Mod. Phys.* **80**, 1215–1274 (2008).
- Zwinger, W. (ed.) *The BCS-BEC Crossover and the Unitary Fermi Gas* (Lecture Notes in Physics Vol. 836, Springer, 2011).
- Frantzeskakis, D. J. Dark solitons in atomic Bose-Einstein condensates: from theory to experiments. *J. Phys. A* **43**, 213001 (2010).
- Burger, S., Bongs, K., Dettmer, S., Erntner, W. & Sengstock, K. Dark solitons in Bose-Einstein condensates. *Phys. Rev. Lett.* **83**, 5198–5201 (1999).
- Denschlag, J. et al. Generating solitons by phase engineering of a Bose-Einstein condensate. *Science* **287**, 97–101 (2000).
- Anderson, B. P. et al. Watching dark solitons decay into vortex rings in a Bose-Einstein condensate. *Phys. Rev. Lett.* **86**, 2926–2929 (2001).
- Becker, C. et al. Oscillations and interactions of dark and dark-bright solitons in Bose-Einstein condensates. *Nature Phys.* **4**, 496–501 (2008).
- Dutton, Z., Budde, M., Slowe, C. & Hau, L. V. Observation of quantum shock waves created with ultra-compressed slow light pulses in a Bose-Einstein condensate. *Science* **293**, 663–668 (2001).
- Engels, P. & Atherton, C. Stationary and nonstationary fluid flow of a Bose-Einstein condensate through a penetrable barrier. *Phys. Rev. Lett.* **99**, 160405 (2007).
- Weller, A. et al. Experimental observation of oscillating and interacting matter wave dark solitons. *Phys. Rev. Lett.* **101**, 130401 (2008).
- Dziarmaga, J. & Sacha, K. Gap soliton in superfluid Fermi gas at zero and finite temperature. *Laser Phys.* **15**, 674–678 (2005).
- Antezza, M., Dalfó, F., Pitaevskii, L. P. & Stringari, S. Dark solitons in a superfluid Fermi gas. *Phys. Rev. A* **76**, 043610 (2007).
- Busch, T. & Anglin, J. R. Motion of dark solitons in trapped Bose-Einstein condensates. *Phys. Rev. Lett.* **84**, 2298–2301 (2000).
- Dziarmaga, J., Karkuszewski, Z. P. & Sacha, K. Quantum depletion of an excited condensate. *Phys. Rev. A* **66**, 043615 (2002).
- Dziarmaga, J. & Sacha, K. Depletion of the dark soliton: the anomalous mode of the Bogoliubov theory. *Phys. Rev. A* **66**, 043620 (2002).
- Law, C. K., Leung, P. T. & Chu, M. C. Quantum fluctuations of coupled dark solitons in a trapped Bose-Einstein condensate. *J. Phys. At. Mol. Opt. Phys.* **35**, 3583–3590 (2002).
- Martin, A. D. & Ruostekoski, J. Quantum and thermal effects of dark solitons in a one-dimensional Bose gas. *Phys. Rev. Lett.* **104**, 194102 (2010).
- Mishmash, R. V. & Carr, L. D. Quantum entangled dark solitons formed by ultracold atoms in optical lattices. *Phys. Rev. Lett.* **103**, 140403 (2009).
- Scott, R. G., Dalfó, F., Pitaevskii, L. P. & Stringari, S. Dynamics of dark solitons in a trapped superfluid Fermi gas. *Phys. Rev. Lett.* **106**, 185301 (2011).
- Spuntarelli, A., Carr, L. D., Pieri, P. & Stringari, G. C. Gray solitons in a strongly interacting superfluid Fermi gas. *New J. Phys.* **13**, 035010 (2011).
- Scott, R. G. et al. The decay and collisions of dark solitons in superfluid Fermi gases. *New J. Phys.* **14**, 023044 (2012).
- Spuntarelli, A., Pieri, P. & Stringari, G. C. Solution of the Bogoliubov–de Gennes equations at zero temperature throughout the BCS-BEC crossover: Josephson and related effects. *Phys. Rep.* **488**, 111–167 (2010).
- Fedichev, P. O., Muryshv, A. E. & Shlyapnikov, G. V. Dissipative dynamics of a kink state in a Bose-condensed gas. *Phys. Rev. A* **60**, 3220–3224 (1999).
- Muryshv, A. E., van Linden van den Heuvell, H. B. & Shlyapnikov, G. V. Stability of standing matter waves in a trap. *Phys. Rev. A* **60**, R2665–R2668 (1999).
- Feder, D. L., Pindzola, M. S., Collins, L. A., Schneider, B. I. & Clark, C. W. Dark-soliton states of Bose-Einstein condensates in anisotropic traps. *Phys. Rev. A* **62**, 053606 (2000).
- Walczak, P. B. & Anglin, J. R. Back-reaction of perturbation wave packets on gray solitons. *Phys. Rev. A* **86**, 013611 (2012).
- Zwierlein, M. W., Abo-Shaeer, J. R., Schirotzek, A., Schunck, C. H. & Ketterle, W. Vortices and superfluidity in a strongly interacting Fermi gas. *Nature* **435**, 1047–1051 (2005).
- Konotop, V. V. & Pitaevskii, L. Landau dynamics of a grey soliton in a trapped condensate. *Phys. Rev. Lett.* **93**, 240403 (2004).
- Liao, R. & Brand, J. Traveling dark solitons in superfluid Fermi gases. *Phys. Rev. A* **83**, 041604 (2011).
- Bertina, G., Pitaevskii, L. & Stringari, S. First and second sound in cylindrically trapped gases. *Phys. Rev. Lett.* **105**, 150402 (2010).
- Keränen, V., Keski-Vakkuri, E., Nowling, S. & Yogendran, K. P. Dark solitons in holographic superfluids. *Phys. Rev. D* **80**, 121901 (2009).
- Keränen, V., Keski-Vakkuri, E., Nowling, S. & Yogendran, K. P. Inhomogeneous structures in holographic superfluids. I. Dark solitons. *Phys. Rev. D* **81**, 126011 (2010).
- Adams, A., Lincoln, D. C., Thomas, S., Peter, S. & John, E. T. Strongly correlated quantum fluids: ultracold quantum gases, quantum chromodynamic plasmas and holographic duality. *New J. Phys.* **14**, 115009 (2012).
- Bulgac, A., Luo, Y.-L. & Roche, K. J. Quantum shock waves and domain walls in the real-time dynamics of a superfluid unitary Fermi gas. *Phys. Rev. Lett.* **108**, 150401 (2012).
- Busch, T. & Anglin, J. R. Dark-bright solitons in inhomogeneous Bose-Einstein condensates. *Phys. Rev. Lett.* **87**, 010401 (2001).
- Middelkamp, S. et al. Dynamics of dark-bright solitons in cigar-shaped Bose-Einstein condensates. *Phys. Lett. A* **375**, 642–646 (2011).
- Karpiuk, T. et al. Spontaneous solitons in the thermal equilibrium of a quasi-1d Bose gas. *Phys. Rev. Lett.* **109**, 205302 (2012).
- Schirotzek, A., Shin, Y., Schunck, C. H. & Ketterle, W. Determination of the superfluid gap in atomic Fermi gases by quasiparticle spectroscopy. *Phys. Rev. Lett.* **101**, 140403 (2008).
- Lutchyn, R. M., Dzero, M. & Yakovenko, V. M. Spectroscopy of the soliton lattice formation in quasi-one-dimensional fermionic superfluids with population imbalance. *Phys. Rev. A* **84**, 033609 (2011).
- Yoshida, N. & Yip, S. K. Larkin-Ovchinnikov state in resonant Fermi gas. *Phys. Rev. A* **75**, 063601 (2007).
- Bulgac, A. & Forbes, M. M. Unitary Fermi supersolid: the Larkin-Ovchinnikov phase. *Phys. Rev. Lett.* **101**, 215301–215304 (2008).
- Radzihovsky, L. Fluctuations and phase transitions in Larkin-Ovchinnikov liquid-crystal states of a population-imbalanced resonant Fermi gas. *Phys. Rev. A* **84**, 023611 (2011).
- Ku, M. J. H., Sommer, A. T., Cheuk, L. W. & Zwierlein, M. W. Revealing the superfluid lambda transition in the universal thermodynamics of a unitary Fermi gas. *Science* **335**, 563–567 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank L. Pitaevskii, S. Stringari, F. Dalfó, W. Zwinger and D. Huse for discussions. This work was supported by the NSF, the ARO MURI on Atomtronics, AFOSR PECASE, ONR, a grant from the Army Research Office with funding from the DARPA OLE programme, and the David and Lucile Packard Foundation.

Author Contributions T.Y. and A.T.S. contributed equally to the experimental work. All authors contributed extensively to the work presented in this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.W.Z. (zwierlein@mit.edu).

Solving the Martian meteorite age conundrum using micro-baddeleyite and launch-generated zircon

D. E. Moser¹, K. R. Chamberlain², K. T. Tait³, A. K. Schmitt⁴, J. R. Darling^{1†}, I. R. Barker¹ & B. C. Hyde³

Invaluable records of planetary dynamics and evolution can be recovered from the geochemical systematics of single meteorites¹. However, the interpreted ages of the ejected igneous crust of Mars differ by up to four billion years^{1–6}, a conundrum⁷ due in part to the difficulty of using geochemistry alone to distinguish between the ages of formation and the ages of the impact events that launched debris towards Earth. Here we solve the conundrum by combining *in situ* electron-beam nanostructural analyses and U–Pb (uranium–lead) isotopic measurements of the resistant micromineral baddeleyite (ZrO₂) and host igneous minerals in the highly shock-metamorphosed shergottite Northwest Africa 5298 (ref. 8), which is a basaltic Martian meteorite. We establish that the micro-baddeleyite grains pre-date the launch event because they are shocked, cogenetic with host igneous minerals, and preserve primary igneous growth zoning. The grains least affected by shock disturbance, and which are rich in radiogenic Pb, date the basalt crystallization near the Martian surface to 187 ± 33 million years before present. Primitive, non-radiogenic Pb isotope compositions of the host minerals, common to most shergottites^{1–4}, do not help us to date the meteorite, instead indicating a magma source region that was fractionated more than four billion years ago^{9–12} to form a persistent reservoir so far unique to Mars^{1,9}. Local impact melting during ejection from Mars less than 22 ± 2 million years ago caused the growth of unshocked, launch-generated zircon and the partial disturbance of baddeleyite dates. We can thus confirm the presence of ancient, non-convecting mantle beneath young volcanic Mars, place an upper bound on the interplanetary travel time of the ejected Martian crust, and validate a new approach to the geochronology of the inner Solar System.

A puzzle in Martian science is that basaltic shergottites yield whole-rock and non-radiogenic mineral Pb–Pb isotopic compositions consistent with a primary Noachian (approximately four billion years (Gyr) ago) age for crystallization⁴, whereas mineral isochrons consistently return ‘young’ late-Amazonian (<0.6 Gyr) ages^{1–3}. Resolution of this shergottite ‘age conundrum’⁷ by U–Pb dating of microcrystals of the accessory phase baddeleyite (ZrO₂) has been pursued^{5,6}, although the small target sizes pose technical challenges. Questions persist as to whether the young dates reflect magmatism or disturbance by metamorphic events on Mars, including shock metamorphism before and during meteorite ejection^{4,7}. Shock experiments aimed at resolving the conundrum predict that shock unloading and the attendant heating do not cause significant Pb-loss from baddeleyite, although it is acknowledged that experiment shock loading paths may not be entirely analogous to nature¹³. Similar experiment-based predictions were made for the mineral zircon, but these did not apply to natural samples where shock heating and related fluid activity were significant^{14,15}. Here we combine *in situ* U–Pb and Pb–Pb isotopic analyses, including the recently developed technique of secondary ion mass-spectrometry (SIMS) micro-baddeleyite dating^{16,17}, with chemical and microstructure data obtained with electron nanobeam techniques including cathodoluminescence and

electron backscatter diffraction (EBSD) that can reveal baddeleyite paragenesis and degree of alteration by shock processes¹⁵ before dating. We have focused on micro-baddeleyites from shergottite NWA 5298 (ref. 8), accession number M53387 from the Royal Ontario Museum’s meteorite collection, one of the most highly shocked shergottites, to demonstrate our methodology for resolving primary and shock-related age records.

Meteorite NWA 5298 is an enriched basaltic (diabasic) shergottite exhibiting a primary igneous texture, which, apart from its relatively oxidized state, is chemically similar to other basaltic shergottites including Shergotty, Zagami and Los Angeles⁸. It is an unbrecciated, highly shocked igneous rock that probably crystallized from a thick lava flow. Partial breakdown of pyroxferroite to fayalite suggests reheating by a later event such as overlying lava flow or shock heating⁸.

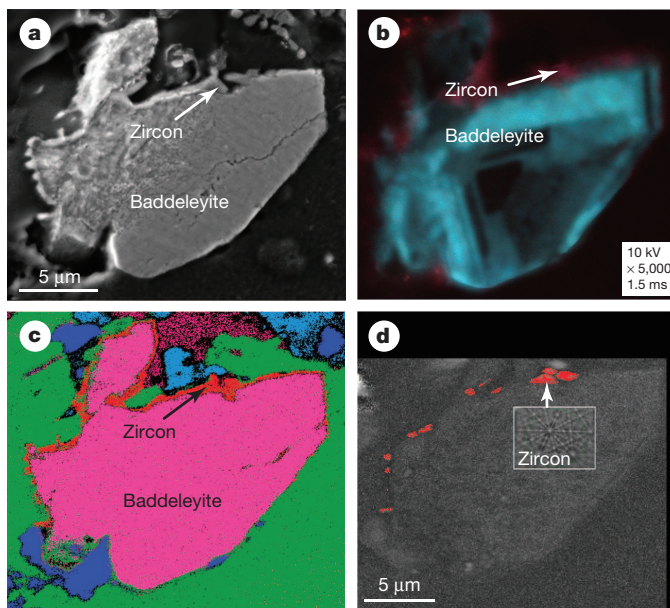


Figure 1 | Nanostructural data for Martian baddeleyite grains showing igneous growth zoning, shock state and launch-generated reaction rim of unshocked zircon. **a**, Secondary electron micrograph of euheedral micro-baddeleyite Grain 16 after partial (upper left) ablation by SIMS. We note its granular texture, caused by shock metamorphism. **b**, Cathodoluminescence image revealing planar trace-element banding typical of igneous crystallization, and zircon rim (magenta) on the side of the grain nearest to a quenched melt pocket. **c**, False-colour element map based on X-ray energy dispersive spectroscopy showing baddeleyite (pink), zircon rim (red) and host maskelynite (green). **d**, EBSD band contrast (diffraction signal strength) map (50 nm step size) indicating weak diffraction in all phases, owing to shock, except for the unshocked zircon rim crystallites (red) that grew during the launch from Mars (the inset shows a representative electron backscatter diffraction pattern for zircon).

¹Department of Earth Sciences, University of Western Ontario, 1151 Richmond Street, London, Ontario N6A 5B7, Canada. ²Department of Geology and Geophysics, University of Wyoming, 1000 East University Avenue, 3006, Laramie, Wyoming 82071, USA. ³Department of Natural History, Mineralogy, Royal Ontario Museum, Toronto, Ontario M5S 2C6, Canada. ⁴Department of Earth and Space Sciences, UCLA, 595 Charles Young Drive, Los Angeles, California 90095, USA. [†]Present address: School of Earth and Environmental Sciences, University of Portsmouth, Burnaby Road, Portsmouth PO1 3QL, UK.

Pervasive maskelynite indicates shock pressures of at least 29 ± 1 GPa and as high as 80 GPa (ref. 8). Baddeleyite grains range in size from $2 \mu\text{m}$ to $30 \mu\text{m}$ in the longest dimension, and occur as euhedral to subhedral elongate to blocky grains intergrown with, or at the boundaries of, larger main-phase minerals, occasionally near quenched melt pockets (such as Grain 16; see Fig. 1 and Supplementary Figs 1 and 3).

Scanning electron microscopy/cathodoluminescence imaging of submicrometre variations in trace-element chemistry revealed several zoning types. Most grains exhibit narrow, bright blue rims surrounding a dark core of weak, broadly concentric, discontinuous, linear zoning. A subset of grains, often with subhedral to euhedral outlines, exhibits oscillatory planar growth banding that is parallel to straight grain margins or forms triangular sector zoning (Fig. 1b). We report similar sector zoning in igneous micro-baddeleyite from an unshocked terrestrial gabbro sill (Supplementary Fig. 2). Common to the exterior surfaces of many grains is a discontinuous and very narrow rim of zircon, up to a few micrometres in width, identified initially by X-ray energy mapping and confirmed by EBSD (Fig. 1c, d; Supplementary Table 1). In many cases the zircon also fills thin fractures within the micro-baddeleyite. Zircon luminesces in the ultraviolet range and has replaced baddeleyite or grown outward from the baddeleyite grain margin along shock microstructures attributable to post-shock decompression⁷.

Structural alteration of the baddeleyite grains by shock is indicated by high-magnification secondary electron and backscatter electron imaging of grain interiors, which reveals roughly equant granules with diameters of about $0.1 \mu\text{m}$ or less (see Fig. 1a). Electron diffraction patterns are sensitive to lattice disorder, as seen in Apollo lunar zircon¹⁸, and here EBSD mapping at maximum resolution (50 nm) indicates that baddeleyite crystallinity is degraded, and lattice orientation is variable, at length scales below 50 nm . The grains appear nearly as amorphous as the host maskelynite glass (Fig. 1d) generated by the Martian impact and launch

event⁸. The only unshocked mineral identified, other than quenched melt pockets, are the zircon rims that crystallized around shocked micro-baddeleyite. High-magnification backscatter electron and EBSD analysis of the zircon (Fig. 1d; Supplementary Table 1) failed to show any evidence of shock microstructures such as planar features, curvilinear features, shock microtwins¹⁵ or reidite (the high-pressure ZrSiO_4 polymorph). Low-magnification imaging of baddeleyite Grain 16 reveals a spatial correlation between the site of zircon growth and the Si-enriched margin of a quenched melt pocket (Supplementary Fig. 3), the latter a common feature of this strongly shock-heated and rapidly cooled shergottite.

SIMS analysis ($n = 20$; all uncertainties at 2σ confidence level) of 15 grains of partially amorphous baddeleyite, all but one (G2457) rimmed by post-shock zircon, revealed relatively low U abundance (about 50 parts per million, p.p.m., to 300 p.p.m.) compared to the reference standard, and proportions of radiogenic Pb that range widely from 97% to 19% (Supplementary Table 2). The U–Pb data (corrected for common Pb measured in neighbouring minerals) form an array parallel to the Tera–Wasserburg concordia, ranging between 227 ± 18 million years (Myr) and 22 ± 2 Myr in apparent age (Fig. 2, Supplementary Table 2), indicative of varying degrees of recent Pb loss. A broad correlation exists between microstructure and $^{206}\text{Pb}/^{238}\text{U}$ dates in that the four grains with preserved igneous zoning are among the six oldest baddeleyite dates, with the oldest grain (G2457), which is unzoned and intergrown with pyroxene, lacking a reaction rim of post-shock zircon. Together, these data are characteristic of an igneous population of uniform age which has experienced variable degrees of recrystallization, reaction to zircon, and single-stage Pb loss during locally intense shock metamorphism.

Our best estimate of the magmatic crystallization age is obtained from the three grains which yielded both the oldest and least disturbed $^{206}\text{Pb}/^{238}\text{U}$ dates (169 ± 16 Myr to 227 ± 18 Myr) and also the highest

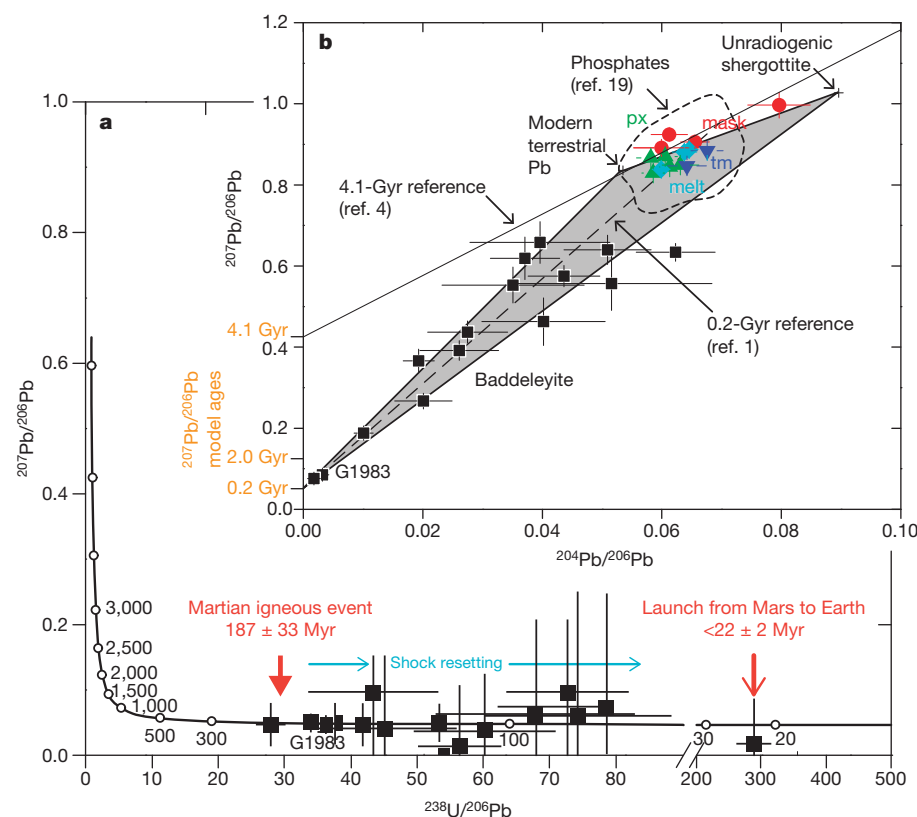


Figure 2 | U–Pb and Pb–Pb (inset) plots of isotopic data (2σ confidence level) for baddeleyite and neighbouring igneous minerals indicating a young crystallization age, variable age disturbance by shock, and inherited primitive Martian Pb. a, Tera–Wasserburg concordia plot of *in situ* SIMS U–Pb baddeleyite data ($n = 20$), corrected for common Pb (see Methods). The original age of the population is interpreted to be 187 ± 33 Myr, based on the weighted $^{206}\text{Pb}/^{238}\text{U}$ dates of the four oldest, most radiogenic analyses, such as G1983 (Supplementary Table 2), with younger grains exhibiting partial Pb loss owing to shock less than 22 ± 2 Myr ago. b, Pb–Pb plot of uncorrected Pb measurements ($n = 15$) from the same baddeleyites (see G1983) compared with SIMS measurements of neighbouring phases (mask, maskelynite; px, pyroxene; tm, titanomagnetite; melt, quenched melt pocket). The baddeleyite Pb–Pb isotopic ratios are insensitive to Pb loss during shock, and are clearly incompatible with a 4.1-Gyr crystallization age, but overlap the 0.2-Gyr isochron originally determined for Shergotty (dashed line)¹. $^{207}\text{Pb}/^{206}\text{Pb}$ model ages are given on the y axis. The scalene triangular distribution reflects mixing of radiogenic, modern terrestrial Pb and a primitive Martian Pb (the reference non-radiogenic Pb composition for depleted shergottite²⁰ is shown). The SIMS common Pb isotopic data for igneous minerals are consistent with arrays of solution and phosphate Pb determined by others^{4,19}, and are a signature of mixing of modern terrestrial Pb with Pb from an ancient U-depleted reservoir unique to Mars^{1,9} that was inherited by the 187 ± 33 -Myr-old magma and NWA 5298.

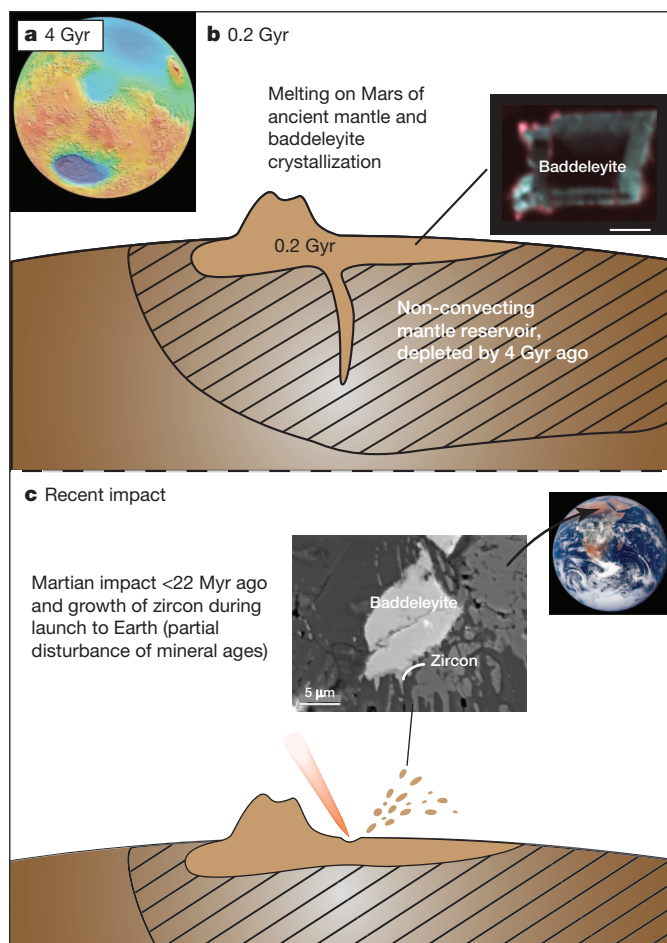


Figure 3 | Three-stage evolution of Martian crust recorded by basaltic shergottite NWA 5298, based on geochemical and nanostructural observations. **a**, Digital elevation map of Mars showing the hemispheric dichotomy caused by an early giant impact²² about 4 Gyr ago, which is one of the processes that may have preserved extremely depleted, ancient mantle reservoirs. **b**, An igneous event at 187 ± 33 Myr (0.2 Gyr) sourced by non-convecting, depleted mantle, most probably due to endogenic hotspot volcanism (for example, Tharsis Montes). The inset shows igneous cathodoluminescence zonation in micro-baddeleyite in the region of NWA 5298 least affected by the later shock event (scale bar, 5 μ m). **c**, Spallation of parent body of meteorite NWA 5298 and other ejecta by oblique impact on Mars at some time after 22 ± 2 Myr ago. The second inset shows ejectic zircon filling shock microstructures (scale bar, 5 μ m) owing to shock-triggered Si release during cooling and the launch toward Earth (Supplementary Fig. 2).

proportions of radiogenic Pb (68% to 96%), such that the $^{206}\text{Pb}/^{238}\text{U}$ dates are least affected by uncertainty in the common Pb correction. On the basis of the weighted mean of the $^{206}\text{Pb}/^{238}\text{U}$ dates ($n = 4$), the best interpretation of the age of baddeleyite and Martian magma crystallization is 187 ± 33 Myr (Fig. 2a). This is consistent with many mineral isochrons from previous studies such as the value of about 200 Myr obtained from high-precision measurements for Shergotty, Zagami and EETA 79001 (ref. 1) (Fig. 2b) and more recent, less precise laser and ion probe results for phosphates or baddeleyite in other shergottites^{5,6,19}. The Pb–Pb data for baddeleyite have a scalene triangular distribution (Fig. 2b) that indicates three-component mixtures of radiogenic Pb with terrestrial common Pb and Martian common Pb from surrounding mineral phases and surfaces. The y -axis intercept of this distribution corresponds to $^{207}\text{Pb}/^{206}\text{Pb}$ model ages of 419 Myr and 0 Myr if regressed through fixed end-member compositions of the most non-radiogenic known Martian Pb (ref. 20) and modern terrestrial Pb, respectively (Fig. 2b). The associated model age uncertainties are comparatively large (hundreds of millions of years) with respect to

the U–Pb geochronology, yet the data are useful for resolving the age conundrum in that they are insensitive to shock-induced isotopic disturbance, and are clearly impossible with regard to a 4-Gyr crystallization age (Fig. 2b).

The isotopic values of common Pb measured in igneous phases (Supplementary Table 3) are also explicable as mixtures of modern terrestrial Pb and primitive Martian Pb (Fig. 2b). Values for pyroxene, maskelynite and titanomagnetite, as well as a melt pocket, overlap the linear array defined by mineral and whole-rock Pb–Pb values from all other basaltic shergottites^{4,19,20}, with the least radiogenic Pb value for maskelynite approaching the most primitive known shergottite value²⁰. The slope of the array therefore has no age significance²⁰ except that the primitive Pb composition signifies an ancient fractionation event in the first several hundred million years of Martian evolution^{9,21}. The primitive Pb of NWA 5298 and other shergottites is therefore best explained as a chemical signature inherited from the upper mantle that melted to form the shergottite parent magma, rather than the age of magma crystallization itself²⁰.

Many workers have pointed out that parts of the upper mantle of Mars must have maintained very low $^{238}\text{U}/^{204}\text{Pb}$ values^{1,3,9,20}, and been chemically isolated—resisting convective mixing since early in Martian evolution. Proposed causes of extreme fractionation of early Mars range from crystallization of an early magma ocean to early partial melting of the planet²¹, perhaps influenced by later planet-scale impact melting²² (Fig. 3a). An ancient source for shergottite magmas is further supported by the preservation of early geochemical signatures such as their $^{182}\text{W}/^{184}\text{W}$, $^{142}\text{Nd}/^{144}\text{Nd}$ and $^{187}\text{Re}/^{187}\text{Os}$ isotopic compositions^{10,11,23}. The upper-mantle source region(s) for shergottites²⁴ are therefore ones that have been tectonically static, but punctuated by large volumes of recent, concentrated melting. A good analogue are the flanks of the Tharsis Montes at the equatorial bulge of Mars, among the largest volcanoes in our Solar System²⁵, which are underlain by stationary volcanic super cells that pierce the highly cratered Noachian crust at the edge of the Martian hemispheric dichotomy²⁶ (Fig. 3b). Rayed craters in the Tharsis Montes have been identified by recent remote sensing as candidates for one of the launch sites of basaltic shergottite meteorites²⁷.

The shock metamorphism caused by the launch event resulted in minor to approximately 80% loss of radiogenic Pb from baddeleyite (Fig. 2a), a phenomenon useful for bracketing the age of launch events. Such Pb loss is not anticipated on the basis of shock experiments¹³, which employ time-separated application of static shock and hours of post-shock annealing that, perhaps owing to additional differences in grain size and porosity, do not reproduce the fine-grained shock heating, metasomatic and quench features which we observe. These include the unshocked zircon rims around and within fractures of most of our baddeleyite grains. The absence of the high-pressure polymorph reidite in rim zircon, and the spatial association of zircon with quenched pockets of impact melt (Supplementary Fig. 3) is consistent with genesis from baddeleyite exposed to post-shock silicic fluids in a relatively low-pressure environment—an ideal agent for Pb loss from baddeleyite along high-diffusivity pathways created by shock microstructures.

The Martian origin for shergottites was first suggested on the basis of the signature of Martian atmospheric chemistry trapped in quenched melt pockets²⁸, here genetically associated with meteorite ejection and zircon growth, and it is reasonable to conclude that the zircon grew as the Martian fragment was cooling and exiting the atmosphere (Fig. 3c). Assuming a present atmosphere thickness of 10 km and a typical ejection velocity of 5 km s^{−1} (ref. 29), the shergottite sample would have experienced localized temperature changes of up to around 2,000 °C (ref. 7) in seconds, with some variation in cooling rate depending on the original size and thermal mass of the ejected body. The rapid pace of events in this scenario is consistent with the local preservation of primary trace-element (cathodoluminescence) zoning in baddeleyites, as well as the extremely thin, micrometre-scale growth of unshocked zircon along shock microstructures (Fig. 3c) which we term “ejectic

zircon". The 22 ± 2 Myr age of the most shock-reset baddeleyite (G1) is therefore an upper limit on launch age and travel time to Earth, complementing existing estimates from cosmic ray exposure dating (see ref. 2 and references therein).

Our integrated microstructural and isotopic analyses allow us to assign specific geologic events to the three stages of Martian geochemical evolution recognized in meteorites since the earliest studies: chemical fractionation of the mantle over 4 Gyr ago, relatively recent melting of that mantle 187 ± 33 Myr ago to cause late-Amazonian basalt magmatism, and a recent launch earthward, perhaps from the flanks of volcanic systems such as Tharsis Montes, triggered by an impact on Mars. Future application of our methodology to Martian meteorites and the broader family of planetary and lunar achondrites promises a more accurate understanding of the inherent, invaluable mineral records of inner Solar System evolution.

METHODS SUMMARY

Micro-baddeleyite grains were located in a petrographic thin section of NWA 5298 by electron beam X-ray mapping. Two subsections were cast in separate epoxy mounts (ROM4 and ROM4b) along with the Phalaborwa baddeleyite standard (2,060 Myr old)³⁰. Electron nanobeam analyses using secondary electron, backscatter electron, cathodoluminescence and EBSD detector systems (Fig. 1) were performed at the University of Western Ontario's Zircon and Accessory Phase Laboratory (ZAPLab) using a Hitachi SU-6600 variable-pressure, field-emission-gun scanning electron microscope according to protocols described elsewhere¹⁵. Electron beam analyses preceded ROM4b SIMS measurements. U–Pb and Pb–Pb *in situ* measurements (ions collected from domains as small as 4 μm) were made with the University of California Los Angeles' CAMECA ims1270 ion microprobe using protocols previously described¹⁷. The proportion of common Pb in NWA 5298 baddeleyite was far higher than that in standards within the same mount (Supplementary Table 2), and in terrestrial and lunar micro-baddeleyite samples measured in the same SIMS session. The elevated common Pb is therefore intrinsic to the Martian sample. The 'blank' common Pb isotopic composition of the baddeleyite was determined from SIMS analysis of neighbouring main-phase minerals and a melt pocket (Supplementary Table 3), and a ²⁰⁸Pb-based method (see Methods) was used to correct for common Pb and determine radiogenic U–Pb ratios (Supplementary Table 2). The determination of the common Pb composition of main-phase minerals by SIMS is described in the Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 27 January; accepted 24 May 2013.

- Chen, J. H. & Wasserburg, G. J. Formation ages and evolution of Shergotty and its parent planet from U–Th–Pb systematics. *Geochim. Cosmochim. Acta* **50**, 955–968 (1986).
- Nyquist, L. E. *et al.* Ages and geologic histories of martian meteorites. *Space Sci. Rev.* **96**, 105–164 (2001).
- Borg, L. & Drake, M. J. A review of meteorite evidence for the timing of magmatism and of surface or near-surface liquid water on Mars. *J. Geophys. Res.* **110**, E12S03 (2005).
- Bouvier, A., Blichert-Toft, J., Vervoort, J. D., Gillet, P. & Albarède, F. The case for old basaltic shergottites. *Earth Planet. Sci. Lett.* **266**, 105–124 (2008).
- Herd, C. D. K., Simonetti, A. & Peterson, N. D. *In situ* U–Pb geochronology of martian baddeleyite by laser ablation MC-ICP-MS. *Lunar Planet. Sci. Conf.* **XXXVIII**, abstr. 1664 (2007).
- Niihara, T. Uranium–lead age of baddeleyite in shergottite Roberts Massif 04261: implications for magmatic activity on Mars. *J. Geophys. Res.* **116**, E12008 (2011).
- El Goresy, A. *et al.* Shock-induced deformation of Shergottites: shock–pressures and perturbations of magmatic ages on Mars. *Geochim. Cosmochim. Acta* **101**, 233–262 (2013).
- Hui, H. *et al.* Petrogenesis of basaltic shergottite Northwest Africa 5298: closed-system crystallization of an oxidized mafic melt. *Meteorit. Planet. Sci.* **46**, 1313–1328 (2011).
- Jagoutz, E. Chronology of SNC meteorites. *Space Sci. Rev.* **56**, 13–22 (1991).
- Brandon, A. D., Walker, R. J., Morgan, J. W. & Goles, G. G. Re–Os isotopic evidence for early differentiation of the martian mantle. *Geochim. Cosmochim. Acta* **64**, 4083–4095 (2000).

- Foley, C. N. *et al.* The early differentiation history of Mars from ¹⁸²W–¹⁴²Nd isotope systematics in the SNC meteorites. *Geochim. Cosmochim. Acta* **69**, 4557–4571 (2005).
- Lapen, T. J. *et al.* A younger age for ALH84001 and its geochemical link to shergottite sources in Mars. *Science* **328**, 347–351 (2010).
- Niihara, T., Kaiden, H., Misawa, K., Sekine, T. & Mikouchi, T. U–Pb isotopic systematics of shock-loaded and annealed baddeleyite: implications for crystallization ages of Martian meteorite shergottites. *Earth Planet. Sci. Lett.* **341–344**, 195–210 (2012).
- Krogh, T. E., Kamo, S. L. & Bohor, B. F. Shock metamorphosed zircons with correlated U–Pb discordance and melt rocks with concordant protolith ages indicate an impact origin for the Sudbury Structure. *Geophys. Monogr.* **95**, 343–353 (1996).
- Moser, D. E. *et al.* New zircon shock phenomena and their use for dating and reconstruction of large impact structures revealed by electron nanobeam (EBSD, CL, EDS) and isotopic U–Pb and (U–Th)/He analysis of the Vredefort dome. *Can. J. Earth Sci.* **48**, 117–139 (2011).
- Chamberlain, K. R. *et al.* *In situ* U–Pb SIMS (IN-SIMS) micro-baddeleyite dating of mafic rocks: method with examples. *Precamb. Res.* **183**, 379–387 (2010).
- Schmitt, A. K., Chamberlain, K. R., Swapp, S. M. & Harrison, T. M. *In situ* U–Pb dating of micro-baddeleyite by secondary ion mass spectrometry. *Chem. Geol.* **269**, 386–395 (2010).
- Nemchin, A. *et al.* Timing of crystallization of the lunar magma ocean constrained by the oldest zircon. *Nature Geosci.* **2**, 133–136 (2009).
- Sano, Y., Terada, K., Takeno, S., Taylor, L. A. & McSween, H. Y. Ion microprobe uranium–thorium–lead dating of Shergotty phosphates. *Meteorit. Planet. Sci.* **35**, 341–346 (2000).
- Gaffney, A. M., Borg, L. E. & Connelly, J. N. Uranium–lead isotope systematics of Mars inferred from the basaltic shergottite QUE 94201. *Geochim. Cosmochim. Acta* **71**, 5016–5031 (2007).
- Borg, L. E., Edmunson, J. E. & Asmerom, Y. Constraints on the U–Pb isotopic systematics of Mars inferred from a combined U–Pb, Rb–Sr, and Sm–Nd isotopic study of the Martian meteorite Zagami. *Geochim. Cosmochim. Acta* **69**, 5819–5830 (2005).
- Nimmo, F., Hart, S. D., Korycansky, D. G. & Agnor, C. B. Implications of an impact origin for the martian hemispheric dichotomy. *Nature* **453**, 1220–1223 (2008).
- Debaillie, V., Brandon, A. D., Yin, Q. Z. & Jacobsen, B. Coupled ¹⁴²Nd–¹⁴³Nd evidence for a protracted magma ocean on Mars. *Nature* **450**, 525–528 (2007).
- Brandon, A. D. *et al.* Evolution of the martian mantle inferred from the ¹⁸⁷Re–¹⁸⁷Os isotope and highly siderophile element abundance systematics of shergottite meteorites. *Geochim. Cosmochim. Acta* **76**, 206–235 (2012).
- Neukum, G. *et al.* The geologic evolution of Mars: episodicity of resurfacing events and ages from cratering analysis of image data and correlation with radiometric ages of Martian meteorites. *Earth Planet. Sci. Lett.* **294**, 204–222 (2010).
- Johnson, C. L. & Phillips, R. J. Evolution of the Tharsis region of Mars: insights from magnetic field observations. *Earth Planet. Sci. Lett.* **230**, 241–254 (2005).
- Tornabene, L. L. *et al.* Identification of large (2–10 km) rayed craters on Mars in THEMIS thermal infrared images: implications for possible Martian meteorite source regions. *J. Geophys. Res.* **111**, E10006 (2006).
- Bogard, D. D. & Johnson, P. Martian gases in an Antarctic meteorite? *Science* **221**, 651–654 (1983).
- Head, J. N., Melosh, H. J. & Ivanov, B. A. Martian meteorite launch: high-speed ejecta from small craters. *Science* **298**, 1752–1756 (2002).
- Heaman, L. M. The application of U–Pb geochronology to mafic, ultramafic and alkaline rocks: an evaluation of three mineral standards. *Chem. Geol.* **261**, 43–52 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge the generosity of NWA 5298 donor D. Gregory. This project was supported by NSERC Discovery Grants to D.E.M. and K.T.T., a Wyoming NASA Space grant to K.R.C., an NSF EAR/IF grant to the UCLA SIMS laboratory, and postdoctoral funding to J.R.D. from the Government of Canada and the University of Western Ontario's Center for Planetary Science and Exploration. We thank S. Swapp and N. Swoboda-Colberg for assistance locating and imaging SIMS targets, and I. Craig (University of Western Ontario) for graphics art support. We also thank L. Nyquist and A. Brandon for reviews of the manuscript.

Author Contributions All authors contributed to this work. D.E.M., K.R.C. and K.T.T. designed the initial project. All authors conducted portions of either, or both, the fundamental field emission gun–scanning electron microscopy and SIMS data collection. A.K.S., K.R.C., D.E.M. and J.R.D. reduced the isotope data. D.E.M., I.R.B. and J.R.D. reduced the field emission gun–scanning electron microscopy data. D.E.M. wrote the main paper, and all authors discussed the results and commented on the manuscript at all stages.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.E.M. (desmond.moser@uwo.ca).

METHODS

Additional SIMS U–Pb analytical details. Analytical conditions followed those described previously^{16,17}, using a primary beam of about 20- μm diameter and oxygen flooding to enhance Pb ionization. Baddeleyite grain lengths ranged from 8 μm to 32 μm , and widths from 2 μm to 12 μm , with aspect ratios as high as 10:1. The CAMECA ims1270 field aperture was set to achieve effective, approximately square, sampling regions of between 4 μm and 8 μm to minimize contributions from host phases. Pb-isotopic fractionation in SIMS is insignificant relative to the level of precision of our measurements, and the agreement between the young Pb–Pb intercept and U–Pb ages (Fig. 2) alleviates any concerns that beam overlap onto phases other than baddeleyite and zircon could have introduced analytical bias⁴. Post-SIMS imaging confirmed that the ion pits were generally well centred on baddeleyite. There is no correlation between $^{206}\text{Pb}/^{238}\text{U}$ date and grain size. Moreover, there is no correlation between percentage radiogenic Pb for an analysis and its Pb peak intensities normalized to the peak intensity for Zr. If the lower radiogenic Pb contents were due to beam overlap with host minerals, then the ratio of Pb content to Zr content would be expected to increase (less Zr from the baddeleyite, more Pb from the host), but this was not observed. Finally, the analyses with the highest Pb/Zr yielded the oldest ages, and those with the lowest Pb/Zr yielded the youngest ages. This trend supports our interpretation that the range in $^{206}\text{Pb}/^{238}\text{U}$ dates reflects varying degrees of loss of radiogenic Pb.

SIMS common Pb correction of micro-baddeleyite U–Pb ages. The abundance of common Pb detected in the unknowns (baddeleyite) was increased relative to terrestrial standard baddeleyite on the same mounts, which necessitated correction

for common Pb derived from two sources—modern terrestrial Pb from the sample surface and Martian Pb intrinsic to the minerals sputtered during micro-baddeleyite SIMS analysis. A procedural common Pb ‘blank’ correction was determined by analysing the main-phase minerals pyroxene, maskelynite and titanomagnetite in mount ROM4b in the vicinity of the dated baddeleyites, using the same run conditions as for micro-baddeleyite unknowns. Monitor species gold was used to track within-run ratio variations with increasing sputtering depth beneath the coated surface, and Pb in a NIST SRM 610 glass cast in the same mount was measured to gauge relative Pb concentrations. An average Pb isotopic composition among all three phases was then determined, weighted by the intensity of the ^{208}Pb peak. The procedural blank composition for the ^{208}Pb -based correction was determined to be $^{208}\text{Pb}/^{206}\text{Pb} = 2.192$ and $^{208}\text{Pb}/^{207}\text{Pb} = 2.548$ (2σ confidence levels of 1.7% and 1.0%, respectively). The ^{204}Pb peak was measured whenever possible, and runs where peak detection showed a >20 p.p.m. deviation between the centring values for $^{94}\text{Zr}_2\text{O}$ and uranium oxide were excluded, such that $^{204}\text{Pb}/^{206}\text{Pb}$ ratios could not be plotted in Fig. 2b.

SIMS measurement of common Pb isotopic composition of NWA 5298. Within-run Pb isotopic ratios and intensities during measurement of main-phase minerals decreased rapidly along with monitor species gold in the first ten of fifty analytical cycles before approaching plateau values. Hence, these plateau values (Supplementary Table 3) are our closest approximation to the Martian Pb isotopic composition of our sample, and indeed our $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{204}\text{Pb}$ ratios of 15.1 and 33.1, respectively, for the least-radiogenic mineral (titanomagnetite) are within the range of whole-rock values for the suite of enriched shergottites⁴.

Suppression of star formation in the galaxy NGC 253 by a starburst-driven molecular wind

Alberto D. Bolatto¹, Steven R. Warren¹, Adam K. Leroy², Fabian Walter³, Sylvain Veilleux¹, Eve C. Ostriker⁴, Jürgen Ott⁵, Martin Zwaan⁶, David B. Fisher¹, Axel Weiss⁷, Erik Rosolowsky^{8†} & Jacqueline Hodge³

The under-abundance of very massive galaxies^{1,2} in the Universe is frequently attributed to the effect of galactic winds^{3–6}. Although ionized galactic winds are readily observable, most of the expelled mass (that is, the total mass flowing out from the nuclear region) is likely to be in atomic^{7,8} and molecular phases^{9–11} that are cooler than the ionized phases. Expanding molecular shells observed in starburst systems such as NGC 253 (ref. 12) and M 82 (refs 13, 14) may facilitate the entrainment of molecular gas in the wind. Although shell properties are well constrained¹², determining the amount of outflowing gas emerging from such shells and the connection between this gas and the ionized wind requires spatial resolution better than 100 parsecs coupled with sensitivity to a wide range of spatial scales, a combination hitherto not available. Here we report observations of NGC 253, a nearby¹⁵ starburst galaxy (distance ~ 3.4 megaparsecs) known to possess a wind^{16–20}, that trace the cool molecular wind at 50-parsec resolution. At this resolution, the extraplanar molecular gas closely tracks the H α filaments, and it appears to be connected to expanding molecular shells located in the starburst region. These observations allow us to determine that the molecular outflow rate is greater than 3 solar masses per year and probably about 9 solar masses per year. This implies a ratio of mass-outflow rate to star-formation rate of at least 1, and probably ~ 3 , indicating that the starburst-driven wind limits the star-formation activity and the final stellar content.

The ionized wind from NGC 253 emerges from its central ~ 200 pc (Fig. 1), and has a central axis approximately in the plane of the sky (inclination $i \approx 78^\circ$) filling a cone of opening angle 60° , with inclination-corrected outflow speeds¹⁸ of a few hundred kilometres per second. Probably because of obscuration, this central wind region is highly asymmetric, with H α emission predominantly on the approaching (front) side of the outflow. Outflow activity extends to ~ 10 kpc, with lobes of H α , X-ray and ultraviolet emission evident above and below the plane of the galaxy¹⁷. Although it is unclear whether NGC 253 hosts an accreting black hole^{21,22}, the corresponding low-luminosity active galactic nucleus is not energetically dominant. The NGC 253 outflow has the characteristic low velocity of starburst-driven winds²³, and is almost certainly driven purely by star formation. Imaging by the Hubble Space Telescope (HST) of the central region reveals absorption lanes due to dust entrained in the outflow, suggesting that the wind may also carry significant amounts of molecular gas. Imaging of the cold molecular wind itself, however, has not been performed.

We imaged the $^{12}\text{CO } J = 1-0$ transition in the central arcminute of NGC 253 using the Atacama Large Millimeter Array (ALMA) located in northern Chile. These data were combined with a map obtained using a 22-m single-dish radio telescope (Mopra; Supplementary Information). The details of the observation and the processing of the data can be found in Supplementary Information. The resulting data cube has a sensitivity of 6 mJy per beam (~ 54 mK) in a 2.5 km s^{-1} bandwidth and an angular resolution of $\theta \approx 3.2''$. The Rayleigh–Jeans

temperature sensitivity of these data is an order of magnitude better than that of previous ^{12}CO interferometric imaging^{12,24} of NGC 253 at its published resolution. This is not the first molecular wind imaged^{9–11}, but these observations make possible the study of the outflow in unprecedented detail, allowing us to map the spatially resolved structure of the cold phase of the galactic wind using a tracer that can be directly related to its mass.

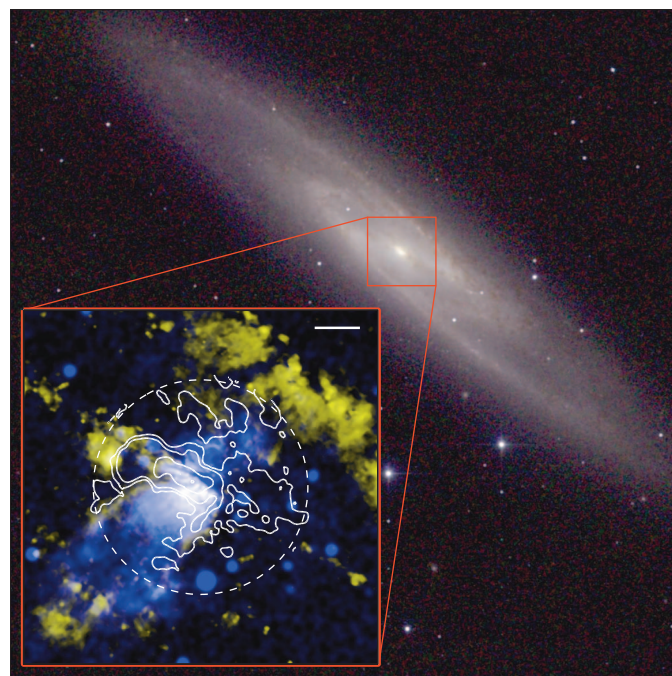


Figure 1 | The warm and hot phases of the galactic wind in NGC 253. The main panel shows the stellar disk of NGC 253 in the Two Micron All Sky Survey (2MASS) JHK composite. Inset, the central 2 kpc of NGC 253 shown in false colour (soft-X-ray emission, blue; H α emission, yellow; image obtained at CTIO and provided by M. Lehnert). The dashed white circle shows the field mapped with ALMA (centred on the reference position: right ascension $\alpha_{2000} = 00 \text{ h } 47 \text{ min } 33.15 \text{ s}$, declination $\delta_{2000} = -25^\circ 17' 17.5''$) with a scale bar of 250 pc in the top right corner. The white contours represent CO emission at $\nu \approx 150\text{--}190 \text{ km s}^{-1}$ and correspond to flux spectral densities $S_\nu = 30, 120, 750$ and $2,500 \text{ mJy per beam}$ (Rayleigh–Jeans brightness temperatures $T_{\text{CO}} \approx 0.25, 1, 7, 22 \text{ K}$), where the extraplanar streamers south of the central bar illustrate the front (approaching) side of the cold molecular wind (see Fig. 2; the contours north of the bar correspond to material at normal rotation velocities in the galaxy disk). The H α emission includes material in the wind as well as emission from normal star-forming regions in the bar and arms of NGC 253. The X-ray image from Chandra¹⁶ shows emission from plasma in the wind as well as point sources in NGC 253 and the background.

¹Department of Astronomy, Laboratory for Millimeter-wave Astronomy, and Joint Space Institute, University of Maryland, College Park, Maryland 20742, USA. ²National Radio Astronomy Observatory, Charlottesville, Virginia 22903, USA. ³Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany. ⁴Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08544, USA. ⁵National Radio Astronomy Observatory, PO Box 0, 1003 Lopezville Road, Socorro, New Mexico 87801, USA. ⁶European Southern Observatory, Karl-Schwarzschild-Strasse 2, 85748 Garching, Germany. ⁷Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany. ⁸Department of Physics and Astronomy, University of British Columbia, Okanagan, Kelowna, British Columbia V1V 1V7, Canada. [†]Present address: Department of Physics, University of Alberta, Edmonton, Alberta T6G 2E1, Canada.

The imaging reveals the presence of previously unknown low-level emission features that are approximately perpendicular to the bright lane of CO emission in the central region of this barred galaxy (Fig. 2). These molecular streamers have surface brightness in the range 30–200 mJy per beam. Perhaps the most prominent feature (which we name SW) is a ridge of CO emission at velocity $v \approx 70$ –250 km s⁻¹ emerging on the southwest edge of the optically bright nuclear region, coincident with a linear dark dust feature along the western edge of the H α and X-ray outflow (Figs 2 and 3). CO emission also extends south from the eastern regions of the nucleus at $v \approx 40$ –140 km s⁻¹, which we associate with the southeast component of the molecular wind and name SE. These features trace an arc almost perfectly aligned with the edges of the H α outflow (Fig. 3), making evident the close spatial correlation between the high-emission-measure ionized filaments and the molecular gas. The receding side of the galactic wind, invisible

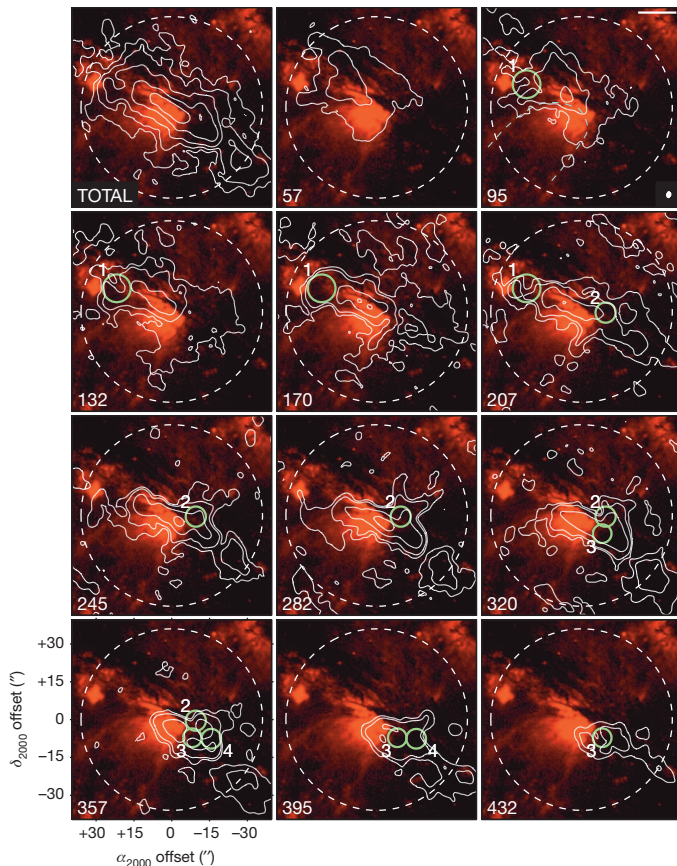


Figure 2 | The cool molecular wind in NGC 253. Maps of ALMA CO(1–0) observations (white contours) against the H α background image (red) showing the outflow (north is up; east is left; reference coordinates are given in Fig. 1; the lower left panel here gives the right ascension and declination offsets). The top left panel shows the intensity integrated over all velocities (contours of $S_{\text{CO}}\Delta v = 5, 10, 25, 100, 250$ Jy km s⁻¹ per beam, where Δv is the velocity interval for the integrated intensity). The other panels are channel maps spanning 37.5 km s⁻¹, with barycentric velocities indicated in the lower left corner of each panel (contours of $S_{\text{CO}} = 30, 120, 750$ and 2,500 mJy per beam). The panels are not corrected by field-of-view illumination, hence the signals are depressed near the edges of the mosaic. The white bar and white filled ellipse in the top right panel represent respectively 250 pc and the synthesized ALMA beam size ($\theta \approx 3.4'' \times 3''$). The dashed circle is the approximate half-power field of view of our seven-pointing mosaic. The green circles numbered 1 to 4 show the location and extent of the four expanding molecular shells identified in the CO cube. The dashed cyan line sketch shown in the channel at 95 km s⁻¹ illustrates the location and shape of the outer bright filaments in the H α outflow¹⁸. The systemic heliocentric velocity of NGC 253 is 243 km s⁻¹, and the emission in the northern regions at 171–208 km s⁻¹ corresponds to material rotating within the disk.

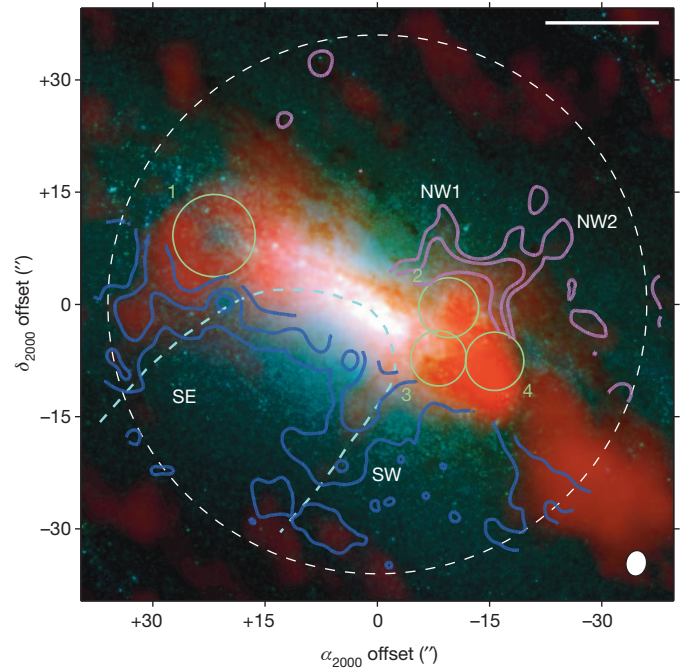


Figure 3 | Integrated molecular wind emission in NGC 253. The background image is a colour composite of HST J-band (shown blue), HST H-band (green), and ALMA CO (red) emission integrated over all velocities and corrected for the mosaic illumination (reference position is given in Fig. 1). The dashed circle, white bar, white filled ellipse and dashed cyan line are as in Fig. 2. The blue and magenta contours show the CO emission in the approaching ($v \approx 73$ –273 km s⁻¹) and receding ($v \approx 208$ –356 km s⁻¹) lobes of the outflow, respectively. The region $\pm 6''$ on each side of the galaxy plane has been blanked for clarity. The contour levels correspond to $S_{\text{CO}}\Delta v = 5, 10$ and 30 Jy km s⁻¹ per beam. The noise in the central regions of the mosaic is ~ 0.3 Jy km s⁻¹ per beam. The green circles illustrate the location and size of the expanding molecular shells.

in H α and barely hinted at in X-ray emission due to absorption from the intervening disk, is apparent in our CO images at velocities $v \approx 240$ –400 km s⁻¹ as a two-pronged structure (that we name NW1 and NW2) located northwest of the nucleus. We see no clear evidence for a corresponding northeast component.

Several of the extraplanar CO features can be traced back in position and velocity to molecular expanding shell structures in the starburst region of NGC 253, providing clues to the launching mechanisms of the molecular wind (Fig. 2). Two of these structures (shells 1 and 3) were found by previous CO observations and dubbed ‘superbubbles’¹². We find four expanding shells, with radii of 60–90 pc, expansion velocities of ~ 23 –42 km s⁻¹, characteristic dynamical ages of ~ 1.4 –4 Myr and molecular masses of $(0.3$ – $1) \times 10^7 M_{\odot}$ (where M_{\odot} is the solar mass; see Supplementary Information). We measure momenta $\sim (8.5$ – $40) \times 10^7 M_{\odot}$ km s⁻¹ and energies $\sim (2$ – $20) \times 10^{52}$ erg. The large momentum involved in each shell suggests driving by the combined effects of multiple stellar winds and (at later stages) supernovae originating in young stellar clusters. These clusters are not directly observed except perhaps for shell 1 (see below). Assuming a steadily driven super-bubble²⁵ with mechanical power $\sim 10^{39}$ erg s⁻¹ for each $10^5 M_{\odot}$ in the driving cluster, from the shell momenta we infer stellar cluster masses of about $(6$ – $40) \times 10^4 M_{\odot}$ if the typical cluster age is $t_{\text{cluster}} \approx 3$ Myr (that is, before supernovae contribute), or lower by a factor $\sim (t_{\text{cluster}}/3 \text{ Myr})^{-1.75}$ for older clusters. Shell 1 corresponds to superbubble SB2¹², and contains a number of sources suggestive of compact clusters in the HST WFC3 Paschen β archival image (not shown). Shell 3 corresponds approximately to superbubble SB1¹², and exhibits [Fe II] 1.644 μ m emission²⁶ coincident with a radio continuum compact source²⁷, presumably a supernova remnant. Shells 2 and 4 are located in regions of very high extinction,

even in the infrared, that contain no sources in the HST images. It is unclear whether shells 3 and 4 are related to features in the wind. Simulations of stellar feedback and galaxy winds find that supernovae, stellar winds and radiation pressure reinforce each other synergistically⁶. We seem to be seeing these mechanisms in action in NGC 253, where the modest expansion velocities of the observed shells probably impart the initial momentum to the molecular gas, which is then advected into the hot outflowing material.

We use the observed CO luminosities and velocities to estimate the mass, mass loss rate and energetics of the molecular wind. The CO luminosity of the four different components identified in the molecular wind (SW, SE, NW1 and NW2) adds up to $\sim 2.0 \times 10^7 \text{ K km s}^{-1} \text{ pc}^2$, approximately equally split between the south and north portions of the outflow. To compute molecular masses in the wind, we adopt an optically thin conversion factor $\alpha_{\text{CO}} \approx 0.34 M_{\odot} \text{ K}^{-1} \text{ km}^{-1} \text{ s pc}^{-2}$, which is about an order of magnitude lower than the value characteristic of the Milky Way disk (Supplementary Information). This results in a total outflowing molecular mass $\sim 6.6 \times 10^6 M_{\odot}$, including the correction for He abundance. The observed projected velocities for the CO wind components are $\sim 30\text{--}60 \text{ km s}^{-1}$ relative to the nearby emission in the nuclear region, with a large de-projection correction due to the low inferred inclination¹⁸ of the outflow (Supplementary Information). These velocities are lower than those of the ionized component¹⁸, suggesting that advection has an important role. The projected filament lengths are $\sim 120\text{--}320 \text{ pc}$, and the implied dynamical filament ages are $\sim 0.3\text{--}1 \text{ Myr}$.

The resulting total molecular mass outflow rate in the wind is $\dot{M}_{\text{w}} \approx 9 M_{\odot} \text{ yr}^{-1}$ with considerable uncertainty due to α_{CO} and the geometrical corrections. Note, however, that most likely corrections would yield an increase in the outflow rate over this value, as our assumptions lead to a conservative estimate (Supplementary Information). Adopting the α_{CO} value used for the central regions of NGC 253 would increase the outflow rate to $\dot{M}_{\text{w}} \sim 30 M_{\odot} \text{ yr}^{-1}$, which we consider a likely upper limit. The star-formation rate in the starburst of NGC 253 is $\sim 2.8 M_{\odot} \text{ yr}^{-1}$, determined from radio continuum and far-infrared measurements²⁸. Obtaining an outflow rate similar to the star-formation rate would require decreasing the CO excitation to $\sim 10 \text{ K}$, increasing the CO abundance by a factor of 3 or increasing the inclination of the outflow to $\sim 45^\circ$, or a combination of the above. These are possible but extreme corrections for a variety of reasons (Supplementary Information). We conclude that $\dot{M}_{\text{w}} \approx 3 M_{\odot} \text{ yr}^{-1}$ is the lower bound of the possible molecular outflow rate values, implying as a robust result that the wind mass loss is at least equal to the star-formation rate and probably a few times higher.

Consequently, the ratio of mass-outflow rate to star-formation rate, η , is certainly ≥ 1 from our lower limit, and probably ~ 3 . Recent estimates based on modelling of unresolved OH absorption spectroscopy¹⁹ suggest $\dot{M}_{\text{w}} \approx 1.6 M_{\odot} \text{ yr}^{-1}$ and possibly as large as $6.4 M_{\odot} \text{ yr}^{-1}$, in approximate agreement with our more direct measurements. More than 90% of the CO luminosity of NGC 253 is found²⁹ within a radius $< 1'$ ($\sim 1 \text{ kpc}$), and, even after taking into account variations in α_{CO} between the disk and the starburst, the majority of the molecular gas is located in this region. Consequently, the central regions of NGC 253 will run out of H_2 in $\sim 60\text{--}120 \text{ Myr}$.

In general, it is not clear what fraction of the outflowing gas actually escapes galaxies, particularly for low-velocity starburst-driven winds like that in NGC 253. Most of the baryons ejected by winds may just linger in the enriched haloes of star-forming galaxies³⁰, to later rain back to their disks, providing fuel for new episodes of star formation. This recycling of baryons constitutes a third mode of galaxy accretion⁵, which may be particularly important in shaping the galaxy mass function at intermediate and large galaxy masses⁴. Our work shows that the mass loading of the starburst-driven wind is substantial, supporting the importance of recycling. With the advent of ALMA, the capabilities for sensitive imaging of the molecular component of galactic winds are

dramatically improving, opening a new window onto the life cycle of baryons in galaxies.

Received 5 March; accepted 4 June 2013.

- Baldry, I. K., Glazebrook, K. & Driver, S. P. On the galaxy stellar mass function, the mass-metallicity relation and the implied baryonic mass function. *Mon. Not. R. Astron. Soc.* **388**, 945–959 (2008).
- Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E. & Hernquist, L. A semi-analytic model for the co-evolution of galaxies, black holes and active galactic nuclei. *Mon. Not. R. Astron. Soc.* **391**, 481–506 (2008).
- Veilleux, S., Cecil, G. & Bland-Hawthorn, J. Galactic winds. *Annu. Rev. Astron. Astrophys.* **43**, 769–826 (2005).
- Davé, R., Oppenheimer, B. D. & Finlator, K. Galaxy evolution in cosmological simulations with outflows—I. Stellar masses and star formation rates. *Mon. Not. R. Astron. Soc.* **415**, 11–31 (2011).
- Oppenheimer, B. D. *et al.* Feedback and recycled wind accretion: assembling the $z = 0$ galaxy mass function. *Mon. Not. R. Astron. Soc.* **406**, 2325–2338 (2010).
- Hopkins, P. F., Quataert, E. & Murray, N. Stellar feedback in galaxies and the origin of galaxy-scale winds. *Mon. Not. R. Astron. Soc.* **421**, 3522–3537 (2012).
- Rupke, D. S., Veilleux, S. & Sanders, D. B. Outflows in active galactic nucleus/starburst-composite ultraluminous infrared galaxies. *Astrophys. J.* **632**, 751–780 (2005).
- Rupke, D. S. N. & Veilleux, S. The multiphase structure and power sources of galactic winds in major mergers. *Astrophys. J.* (in the press); preprint at <http://arxiv.org/abs/1303.6866> (2013).
- Walter, F., Weiss, A. & Scoville, N. Molecular gas in M82: resolving the outflow and streamers. *Astrophys. J.* **580**, L21–L25 (2002).
- Feruglio, C. *et al.* Quasar feedback revealed by giant molecular outflows. *Astron. Astrophys.* **518**, L155–L158 (2010).
- Alatalo, K. *et al.* Discovery of an active galactic nucleus driven molecular outflow in the local early-type galaxy NGC 1266. *Astrophys. J.* **735**, 88–99 (2011).
- Sakamoto, K. *et al.* Molecular superbubbles in the starburst galaxy NGC 253. *Astrophys. J.* **636**, 685–697 (2006).
- Weiss, A., Walter, F., Neininger, N. & Klein, U. Evidence for an expanding molecular superbubble in M 82. *Astron. Astrophys.* **345**, L23–L25 (1999).
- Matsushita, S. *et al.* Starburst at the expanding molecular superbubble in M82: self-induced starburst at the inner edge of the superbubble. *Astrophys. J.* **618**, 712–722 (2005).
- Dalcanton, J. J. *et al.* The ACS nearby galaxy survey treasury. *Astrophys. J. Suppl. Ser.* **183**, 67–108 (2009).
- Strickland, D. K., Heckman, T. M., Weaver, K. A. & Dahlem, M. Chandra observations of NGC 253: new insights into the nature of starburst-driven superwinds. *Astron. J.* **120**, 2965–2974 (2000).
- Strickland, D. K., Heckman, T. M., Weaver, K. A., Hoopes, C. G. & Dahlem, M. Chandra observations of NGC 253. II. On the origin of diffuse X-ray emission in the halos of starburst galaxies. *Astrophys. J.* **568**, 689–716 (2002).
- Westmoquette, M. S., Smith, L. J. & Gallagher, J. S. III Spatially resolved optical integral field unit spectroscopy of the inner superwind of NGC 253. *Mon. Not. R. Astron. Soc.* **414**, 3719–3739 (2011).
- Heckman, T. M., Lehnert, M. D., Strickland, D. K. & Armus, L. Absorption-line probes of gas and dust in galactic superwinds. *Astrophys. J. Suppl. Ser.* **129**, 493–516 (2000).
- Sugai, H., Davies, R. I. & Ward, M. J. The collimated wind in NGC 253. *Astrophys. J.* **584**, L9–L12 (2003).
- Forbes, D. A., Polehampton, E., Stevens, I. R., Brodie, J. P. & Ward, M. J. A multiwavelength view at the heart of the superwind in NGC 253. *Mon. Not. R. Astron. Soc.* **312**, 689–697 (2000).
- Weaver, K. A., Heckman, T. M., Strickland, D. K. & Dahlem, M. Chandra observations of the evolving core of the starburst galaxy NGC 253. *Astrophys. J.* **576**, L19–L23 (2002).
- Sturm, E. *et al.* Massive molecular outflows and negative feedback in ULIRGs observed by Herschel-PACS. *Astrophys. J.* **733**, L16–L20 (2011).
- Sakamoto, K. *et al.* Star-forming cloud complexes in the central molecular zone of NGC 253. *Astrophys. J.* **735**, 19–30 (2011).
- McCray, R. & Kafatos, M. Supershells and propagating star formation. *Astrophys. J.* **317**, 190–196 (1987).
- Alonso-Herrero, A., Rieke, G. H., Rieke, M. J. & Kelly, D. M. The [Fe II] 1.644 micron emission in M82 and NGC 253: is it a measure of the supernova rate? *Astron. J.* **125**, 1210–1225 (2003).
- Ulvestad, J. S. & Antonucci, R. R. J. VLA observations of NGC 253: supernova remnants and H II regions at 1 parsec resolution. *Astrophys. J.* **488**, 621–641 (1997).
- Ott, J., Weiss, A., Henkel, C. & Walter, F. The temperature distribution of dense molecular gas in the center of NGC 253. *Astrophys. J.* **629**, 767–780 (2005).
- Young, J. S. *et al.* The FCRAO extragalactic CO survey. I. The data. *Astrophys. J. Suppl. Ser.* **98**, 219–257 (1995).
- Tumlinson, J. *et al.* The large, oxygen-rich halos of star-forming galaxies are a major reservoir of galactic metals. *Science* **334**, 948–952 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements A.D.B. acknowledges partial support from a CAREER grant NSF-AST0955836, NSF-AST1139998 and from a Research Corporation for Science Advancement Cottrell Scholar award. S.V. acknowledges partial support through grant NSF-AST100958. E.C.O. is supported by the NSF through grant AST-0908185. ALMA is

a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada) and NSC and ASIAA (Taiwan), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. The National Radio Astronomy Observatory is a facility of the NSF operated under cooperative agreement by Associated Universities, Inc. We thank M. Lehnert for providing the H α image, processed by himself and M. Dahlem.

Author Contributions A.D.B. and S.R.W. performed the detailed calculations used in the analysis. A.K.L., S.R.W. and A.D.B. reduced and analysed the ALMA data. D.B.F. reduced and analysed the HST archival data. A.D.B., F.W., A.K.L. and M.Z. wrote the ALMA

proposal and designed the observations with input from co-authors. J.O. obtained and reduced the Mopra observations. A.D.B. wrote the manuscript with input from F.W., A.K.L., S.R.W., S. V. and E.C.O. All authors were participants in the discussion of results, determination of the conclusions and revision of the manuscript.

Author Information This paper makes use of the following ALMA data: ADS/JAO.ALMA#2011.0.00172.S. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.D.B. (bolatto@astro.umd.edu).

Insights into the phylogeny and coding potential of microbial dark matter

Christian Rinke¹, Patrick Schwientek¹, Alexander Sczyrba^{1,2}, Natalia N. Ivanova¹, Iain J. Anderson^{1,‡}, Jan-Fang Cheng¹, Aaron Darling^{3,4}, Stephanie Malfatti¹, Brandon K. Swan⁵, Esther A. Gies⁶, Jeremy A. Dodsworth⁷, Brian P. Hedlund⁷, George Tsiamis⁸, Stefan M. Sievert⁹, Wen-Tso Liu¹⁰, Jonathan A. Eisen³, Steven J. Hallam⁶, Nikos C. Kyrpides¹, Ramunas Stepanauskas⁵, Edward M. Rubin¹, Philip Hugenholtz¹¹ & Tanja Woyke¹

Genome sequencing enhances our understanding of the biological world by providing blueprints for the evolutionary and functional diversity that shapes the biosphere. However, microbial genomes that are currently available are of limited phylogenetic breadth, owing to our historical inability to cultivate most microorganisms in the laboratory. We apply single-cell genomics to target and sequence 201 uncultivated archaeal and bacterial cells from nine diverse habitats belonging to 29 major mostly uncharted branches of the tree of life, so-called ‘microbial dark matter’. With this additional genomic information, we are able to resolve many intra- and inter-phylum-level relationships and to propose two new superphyla. We uncover unexpected metabolic features that extend our understanding of biology and challenge established boundaries between the three domains of life. These include a novel amino acid use for the opal stop codon, an archaeal-type purine synthesis in Bacteria and complete sigma factors in Archaea similar to those in Bacteria. The single-cell genomes also served to phylogenetically anchor up to 20% of metagenomic reads in some habitats, facilitating organism-level interpretation of ecosystem function. This study greatly expands the genomic representation of the tree of life and provides a systematic step towards a better understanding of biological evolution on our planet.

Microorganisms are the most diverse and abundant cellular life forms on Earth, occupying every possible metabolic niche. The large majority of these organisms have not been obtained in pure culture and we have only recently become aware of their presence mainly through cultivation-independent molecular surveys based on conserved marker genes (chiefly small subunit ribosomal RNA; SSU rRNA) or through shotgun sequencing (metagenomics)^{1,2}. As an increasing number of environments are deeply sequenced using next-generation technologies, diversity estimates for Bacteria and Archaea continue to rise, with the number of microbial ‘species’ predicted to reach well into the millions³. According to SSU rRNA-based phylogeny, these fall into at least 60 major lines of descent (phyla or divisions) within the bacterial and archaeal domains⁴, of which half have no cultivated representatives (so-called ‘candidate’ phyla). This biased representation is even more fundamentally skewed when considering that more than 88% of all microbial isolates belong to only four bacterial phyla, the Proteobacteria, Firmicutes, Actinobacteria and Bacteroidetes (Supplementary Fig. 1a). Genome sequencing of microbial isolates naturally reflects this cultivation bias (Supplementary Fig. 1b). Recently, a systematic effort, the Genomic Encyclopaedia of Bacteria and Archaea (GEBA) Project⁵, has been initiated to maximize coverage of the diversity captured in microbial isolates by phylogenetically targeted genome sequencing. However, GEBA does not address candidate phyla that represent a major unexplored portion of microbial diversity, and have been referred to as microbial dark matter (MDM)⁶.

Metagenomics can obtain genome sequences from uncultivated microorganisms through direct sequencing of DNA from the environment⁷.

In some instances, draft or even complete genomes of candidate phyla have been recovered solely from metagenomic data (Supplementary Table 1). A complementary cultivation-independent approach for obtaining genomes from candidate phyla is single-cell genomics; the amplification and sequencing of DNA from single cells obtained directly from environmental samples⁸. This approach can be used for targeted recovery of genomes and has been applied to members of several candidate phyla (Supplementary Table 1). In particular, natural populations that have a high degree of genomic heterogeneity will be more accessible through single-cell genomics than through metagenomics as co-assembly of multiple strains is avoided. Despite these advances in obtaining genomic representation of MDM, no systematic effort has been made to obtain genomes from uncultivated candidate phyla using single-cell whole genome amplification approaches.

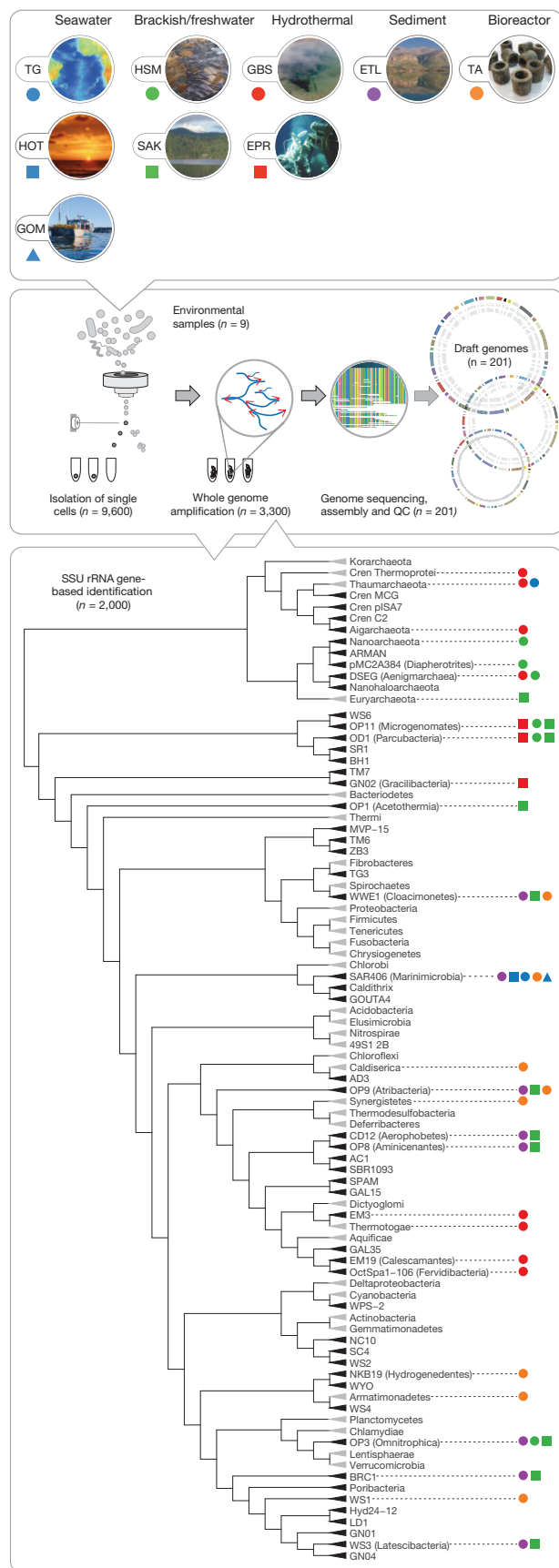
Here, we present GEBA-MDM, the natural extension of the Genomic Encyclopaedia into uncultivated diversity by applying single-cell genomics to recover draft genomes from over 200 cells representing more than 20 major uncultivated archaeal and bacterial lineages. Genome-based phylogenetic analysis confirms the validity of rRNA-defined candidate phyla as monophyletic groups and resolves a number of associations among phyla not apparent by single gene analysis. We discovered several unexpected features, including archaeal sigma factors and stop codon reassignments that challenge established views of the microbial world. Furthermore, we show that single-cell genome references substantially improve the phylogenetic anchoring of about 340 million previously incorrectly or under-classified metagenomic reads.

¹DOE Joint Genome Institute, Walnut Creek, California 94598, USA. ²Center for Biotechnology, Bielefeld University, 33602 Bielefeld, Germany. ³Department of Evolution and Ecology, University of California Davis, Davis, California 95616, USA. ⁴three institute, University of Technology Sydney, Ultimo NSW 2007, Australia. ⁵Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine 04544-0380, USA.

⁶Department of Microbiology and Immunology and Graduate Program in Bioinformatics, University of British Columbia, Vancouver, British Columbia, V6T 1Z3 Canada. ⁷School of Life Sciences, University of Nevada, Las Vegas, Nevada 89154-4004, USA. ⁸Department of Environmental and Natural Resources Management, University of Patras, Agrinio, T.K. 30100, Greece. ⁹Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. ¹⁰Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61802, USA.

¹¹Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences and Institute for Molecular Bioscience, The University of Queensland, St. Lucia QLD 4072, Australia.

[‡]Deceased.



Single-cell genomics at scale

We began by screening numerous physicochemically and geographically diverse environmental samples using SSU rRNA community profiling to identify habitats enriched in candidate phyla, and we targeted nine for in-depth single-cell analysis (Fig. 1, top panel, and Supplementary Fig. 2). Cells representing novel lineages were identified using high-throughput single-cell flow sorting, whole-genome amplification and SSU rRNA screening of single amplified genomes (SAGs; Fig. 1, middle panel; see Methods). A total of 201 SAGs representing 21 and 8 highly under-represented major bacterial and archaeal lineages were selected for whole genome sequencing (Fig. 1, bottom panel).

To improve assemblies, SAG sequence data was digitally normalized to reduce over-represented regions caused by amplification bias⁹. The fidelity of the resulting assemblies was validated using tetra-nucleotide frequency, BLAST (Basic Local Alignment Search Tool) and single copy marker gene analyses (Supplementary Methods and Supplementary Fig. 4). Draft SAGs ranged in size from 148 kilobase pairs (kb) to 2.4 Mb comprising an average of 59 major contigs per assembly (Supplementary Fig. 5a and Supplementary Table 2). Genome completeness was estimated to range from less than 10% to more than 90% (mean 40%) based on the presence or absence of 139 bacterial and 162 archaeal conserved marker genes (Supplementary Fig. 5a). Combining reads of single cells belonging to the same population, that is, with an average nucleotide identity of $\geq 97\%$ (ref. 10) (see Methods), improved assemblies and produced seven population genomes with an estimated completeness of over 90% (Supplementary Fig. 6 and Supplementary Fig. 5a, b).

Genome-based phylogenetic inference

SSU rRNA trees are known to be sound predictors of phylogenetic novelty^{5,11} despite the blurring of vertical descent by lateral gene transfer¹². However, concatenated alignments of multiple universally distributed single copy marker genes are generally considered to provide greater phylogenetic resolution than any individual gene for estimating a species tree¹³. We constructed bootstrapped maximum likelihood trees based on a concatenation of up to 38 commonly used conserved marker genes^{5,14} (Supplementary Methods and Supplementary Table 3) with 15 taxa configurations¹⁵ (Supplementary Table 4). Substitution models were selected to address known issues, including long branch attraction¹⁶ (discussed further in Supplementary Information). Congruency of the individual marker gene topologies to each other was independently assessed confirming the selection of these gene

Figure 1 | Sampling sites and single-cell sequencing workflow. Upper panel, nine global sampling sites grouped into ocean samples (blue), fresh and brackish water samples (green), hydrothermal sites (red), sediment samples (magenta), and bioreactor samples (orange symbol). EPR, East Pacific Rise; ETL, Etoliko Lagoon; GBS, Great Boiling Spring; GOM, Gulf of Maine; HOT, Hawaii Ocean Time-series Project; HSM, Homestake Mine; SAK, Sakinaw Lake; TA, terephthalate degrading reactor; TG, tropical gyre in the South Atlantic. Middle panel, environmental samples were processed using a fluorescence-activated cell sorter allowing the isolation of 9,600 single cells. Each cell was lysed and the genome amplified yielding 3,300 successful amplifications. Resulting SAGs were screened by SSU rRNA gene PCR and sequencing to resolve taxonomic identities. SAGs belonging to major novel lineages were selected for genome sequencing and assembly resulting in 201 draft genomes. QC, quality control. Lower panel, cladogram showing the taxonomy of the SSU rRNA gene sequences, grouped into phyla. Candidate phyla are highlighted in black, and known phyla (according to the list of 'Prokaryote Names with Standing in Nomenclature' at <http://www.bacterio.net/>) are shown in light grey. For each phylum for which we retrieved one or more single-cell genomes the sampling sites are indicated according to the symbols in the upper panel. Note that marker gene phylogeny suggests that SAG JGI000068-E11 clusters within the PER group, a sister lineage to Gracilbacteria (Supplementary Fig. 3). This finding is not supported by the SSU rRNA gene phylogeny and will need further evaluation as more genome and SSU rRNA gene sequences become available.

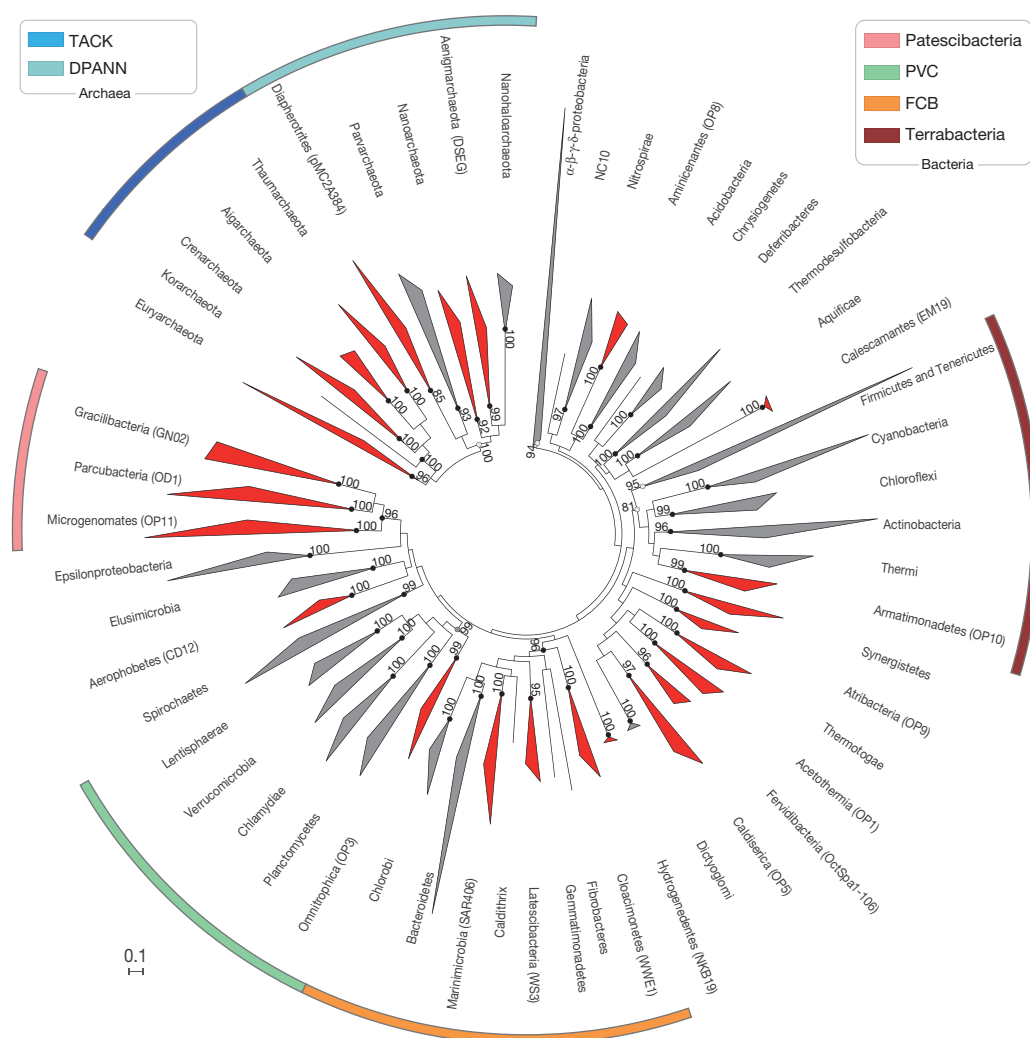


Figure 2 | Maximum-likelihood phylogenetic inference of Archaea and Bacteria. The phylogenetic trees are based on up to 38 marker genes and sequences are collapsed at the phylum level occluding subgroups such as the Geoarchaeota which clusters within the Crenarchaeota. Phyla containing SAGs from this study are highlighted in red. Superphyla (TACK, DPANN, Terrabacteria, FCB, PVC and Patescibacteria) are highlighted with colour ranges. The phylogenetic robustness (monophyly score) of phyla and superphyla is indicated by a small circle on the node: black circle (node was

resolved in 100% of all tree calculations); grey circle (resolved in $\geq 90\%$ of all calculations); light-grey circle (resolved in $\geq 50\%$ of all calculations). Average bootstrap support values are provided for each phylum and superphylum when resolved. The underlying phylogenetic inference configurations as well as detailed branch support values and monophyly scores are provided in Supplementary Table 3. The two domain trees were independently calculated and are unrooted and the scale bar represents 10% estimated sequence divergence for both trees.

families for genome tree reconstruction (Supplementary Fig. 7). All candidate phyla with three or more SAG representatives were resolved as monophyletic groups consistent with their rRNA delineations (Fig. 2 and Supplementary Fig. 8). These are the first substantive genomic data for candidate bacterial phyla SAR406 (Marine Group A)¹⁷, OP3, OP8 (ref. 18), WS1, WS3 (ref. 19), BRC1, CD12, EM19, EM3, NKB19, and Oct-Spa1-106 (ref. 20), as well as for several highly divergent archaeal groups related to the Nanoarchaeota (Fig. 2). We propose names for candidate phyla with two or more representatives based on their inferred physiology and distinguishing properties (Supplementary Table 5, see below).

Owing to the greater phylogenetic resolution afforded by the concatenated gene data sets, compared to rRNA phylogeny, we were able to identify a number of robust associations among phyla. These include the well-recognized Planctomycetes–Verrucomicrobia–Chlamydiae (PVC) superphylum that, based on rRNA analysis, was proposed to also include candidate phylum Omnitrophica (OP3) and the phylum Lentisphaerae²¹. Genome-based analysis confirms this grouping (Fig. 2) and we found a suggested PVC signature gene²² in an Omnitrophica genome (Supplementary Information). The Fibrobacteres–Chlorobi–Bacteroidetes (FCB) superphylum²³ was robustly resolved

together with Marinimicrobia (SAR406), Latescibacteria (WS3), Cloacimonetes (WWE1), Gemmatimonadetes²⁴ and Caldithrix²⁵. Comparative genomics revealed that a conserved carboxy-terminal domain of extracellular proteinases (TIGR04183) is found exclusively (but not comprehensively) in members of the FCB superphylum. This includes the original phyla Fibrobacteres, Chlorobi, Bacteroidetes, as well as the candidate phyla Cloacimonetes, Marinimicrobia, Latescibacteria and the Caldithrix genome (Supplementary Information).

The Terrabacteria, proposed to comprise the ‘terrestrial’ bacterial phyla Actinobacteria, Cyanobacteria, Thermi (Deinococcus–Thermus), Chloroflexi and Firmicutes²⁶, was resolved in our analysis with the additional membership of Armatimonadetes (former candidate phylum OP10)²⁷ (Fig. 2). Perhaps more compelling than the assertion of ancient adaptations to life on land unifying the Terrabacteria²⁶ are commonalities in cell envelope architecture. This superphylum comprises monoderm (single membrane) and atypical monoderm lineages²⁸. We assessed the additional proposed Terrabacteria phyla for genes most characteristic of monoderms and diderms²⁹ and confirmed that all had monoderm-like or atypical gene complements (Supplementary Fig. 9).

The phylogenetic placement of the Cloacimonetes (WWE1 clade) has been inconclusive based on rRNA comparative analysis. It was

originally proposed as a candidate phylum³⁰ and more recently as a class within the Spirochaetes phylum²⁸. Our analysis, which substantially expands the genomic representation of this group, finds no support for a specific affiliation with the Spirochaetes (Fig. 2). It was suggested, based on a smaller data set, that the Acidobacteria reproducibly cluster with the Deltaproteobacteria¹⁴ but this is not supported by our analyses. Instead, Acidobacteria reproducibly affiliate with the Aminacenantales (OP8) (Fig. 2). Candidate phylum OP11, as originally proposed²⁶, has not been resolved consistently as a monophyletic group leading to the proposal for subdivision into multiple phyla, including OP11 (former subdivisions 1 to 3 only), OD1 (former OP11 subdivision 5) and SR1 (ref. 31). Here we found that Microgenomates (OP11) and Parcubacteria (OD1) genomes were resolved reproducibly as a monophyletic group based on concatenated marker gene analysis together with Gracilibacteria (GN02)³². To recognize this affiliation, we propose the superphylum name 'Patescibacteria' (patesco (Latin), meaning bare) (Fig. 2), reflecting the reduced metabolic capacities of these lineages³³. We found support for a specific association between the Patescibacteria and Terrabacteria using a larger bacteria-specific marker gene set (Supplementary Fig. 10). This association is consistent with a monoderm-like gene complement in the Patescibacteria (Supplementary Fig. 9) but will need to be verified when additional genomes belonging to these lineages are available.

Based on phylogenetic analysis of our archaeal single-cell genomes and several recently described genome-sequenced lineages of very small cells, such as *Candidatus* Parvarchaeum, *Candidatus* Micrarchaeum³⁴, *Candidatus* Nanosalina, *Candidatus* Nanosalinarum³⁵, we propose the following phyla: Diapherotrites (pMC2A384)³⁶, Parvarchaeota, Aenigmarchaeota (DSEG)³⁷ and Nanohaloarchaeota (Fig. 2 and Supplementary Table 5). The Nanohaloarchaeota include the recently proposed class Nanohaloarchaea that was incorrectly placed within the Euryarchaeota owing to inadequate outgroup representation³⁵. We predict that small cell and genome size are unifying features of these phyla and in Archaea-only trees these lineages, together with the Nanoarchaeota, form a monophyletic superphylum for which we propose the identifier, DPANN (Fig. 2 and Supplementary Text). Our expanded genomic representation and analysis of the archaeal domain also supports the proposal for the TACK superphylum³⁸, but is not consistent with the eocyte hypothesis³⁹, which places the Eukaryota within the archaeal domain, recently reinvestigated using a 36-genome data set⁴⁰ (Supplementary Fig. 11). As more genomes and improved phylogenetic inference methods come to hand, our proposed lineage delineations can be further evaluated.

Functional diversity and novel findings

The numerous strategies that cultivated microorganisms use to obtain energy and nutrients suggest that many metabolic surprises remain to be discovered in the uncultivated microbial majority. Here we provide a first glimpse into the potential functional diversity of many of the investigated candidate phyla and novel lineages. The majority of bacterial and several archaeal single-cell genomes in our study possess a large array of genes for the degradation of amino acids and sugars (providing the basis for some candidate names for phyla; Supplementary Table 5), pointing to a heterotrophic lifestyle (Supplementary Fig. 12). We found evidence for an electron transport chain, and thus the ability to perform a more complete set of cellular respiration processes, in most bacterial SAGs with the notable exception of members of the Parcubacteria (OD1), Microgenomates (OP11), Gracilibacteria (GN02) and Latescibacteria (WS3). Genes necessary for carbon fixation were found in a wide range of archaeal SAGs (Wood-Ljungdahl pathway, adenosine nucleotide degradation pathway) with a more limited distribution in the bacterial SAGs (Supplementary Fig. 12). Hydrogen metabolism is widespread amongst the novel lineages, and two SAGs (belonging to Caldiserica and Aigarchaeota) have genes for sulphur utilization (Supplementary Fig. 12 and Supplementary Table 6).

A novel recoding of the opal stop codon UGA for glycine was identified in members of the Gracilibacteria (Fig. 3 and Supplementary Fig. 13a). The same recoding was found and biochemically validated in candidate phylum SR1 very recently⁴¹, suggesting that this codon reassignment may be phylogenetically widespread in uncharacterized lineages. This expands the known alternative coding for UGA, which has previously been reported for selenocysteine⁴² and tryptophan^{43,44}. The very low guanine–cytosine content of the Gracilibacteria single-cell genomes (<24%) may have driven the recoding of UGA to a lower guanine–cytosine glycine codon alternative (UGA versus GGN) particularly as glycine is the third most commonly used amino acid (>7% average abundance per genome; Supplementary Fig. 13b).

Purine biosynthesis is highly conserved in the Bacteria and Archaea in terms of the penultimate step in the pathway responsible for ribonucleotide formylation⁴⁵. All bacteria sequenced so far use the PurH1 enzyme for this step, whereas the majority of Archaea use the PurP enzyme. However, members of the bacterial superphylum Patescibacteria lack the *purH1* gene and instead have an euryarchaeal *purP*-like gene (Fig. 3

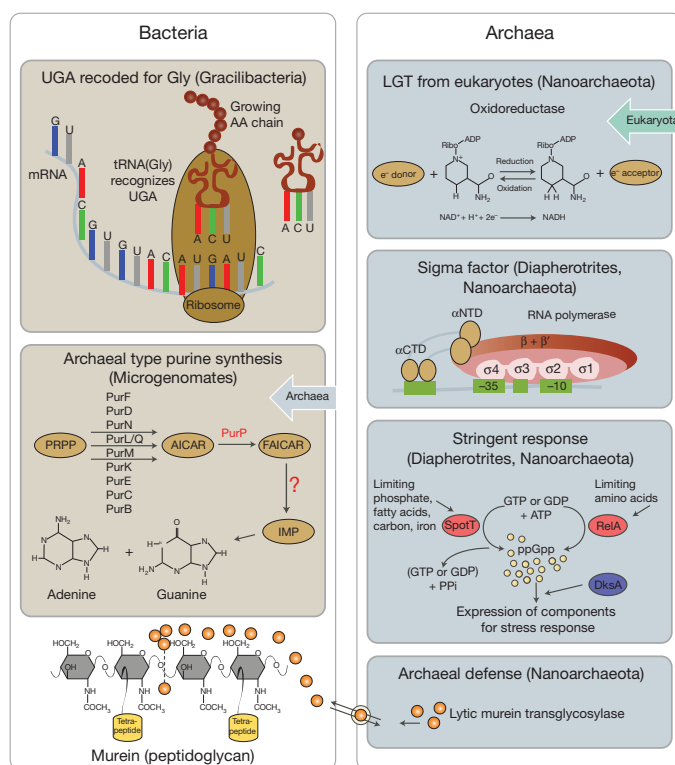


Figure 3 | Novel metabolic features found in the SAG data set. Left, features found in Bacteria: in a subgroup of the Gracilibacteria (GN02), the opal stop codon UGA codes for glycine and these genomes encode a transfer RNA (tRNA) for UGA. Two lineages of Microgenomates (OP11) bacteria use the archaeal pathway (PurH1 enzyme) for purine (adenine, guanine) biosynthesis, inferred to have been acquired by lateral gene transfer (LGT) from Euryarchaeota. AICAR, aminoimidazole carboxamide ribonucleotide; ATP, adenosine tri-phosphate; FAICAR, formyl aminoimidazole carboxamide ribonucleotide; IMP, inosine monophosphate; mRNA, messenger RNA; PRPP, phosphoribosyl pyrophosphate; PurH, bifunctional purine biosynthesis protein PurH. Right, features found in Archaea. A Nanoarchaeota genome encodes an oxidoreductase most closely related to the soil-living amoebae (slime mould) representing a lateral gene transfer from a eukaryote to an archaeon. Two members of the Diapherotrites (pMC2A384) and one representative of the Nanoarchaeota encode complete bacteria-like sigma factors ($\sigma 70$). The bacterial stringent response based on deployment of signalling molecules (ppGpp) was identified in a member of the Diapherotrites and the Nanoarchaeota. A bacterial-like lytic murein transglycosylase was found in two members of the Nanoarchaeota. α CTD, α -subunit C-terminal domain; α NTD, α -subunit N-terminal domain; ADP, adenosine di-phosphate; GDP, guanosine di-phosphate; GTP, guanosine tri-phosphate.

and Supplementary Table 7) as a result of an ancient lateral transfer of most of the purine biosynthesis operon from a Thermococci-like donor to the ancestor of the Patescibacteria (Supplementary Fig. 14).

The DPANN superphylum contains a number of metabolic novelties pointing to a capacity for co-opting foreign genetic elements. A Nanoarchaeota genome encodes an oxidoreductase most closely related to the slime mould *Dictyostelium discoideum* and sits within the eukaryal evolutionary radiation for this gene (Supplementary Fig. 15). To our knowledge, this is the first instance of a lateral gene transfer from a eukaryote to an archaeon. Sigma factors are RNA transcription initiation factors found exclusively in Bacteria, although one conserved sigma factor domain (region four) has been reported in Archaea⁴⁶. Here we report the first instance of complete bacteria-like sigma factors ($\sigma 70$) in Archaea, specifically in two members of the Diapherotrites and one representative of the Nanoarchaeota (Fig. 3 and Supplementary Table 8). These appear to be the result of multiple lateral transfers from bacterial donors (Supplementary Fig. 16). All three sigma factors belong to the non-essential $\sigma 70$ groups (3 and 4)⁴⁷ and their hosts retain the standard archaeal TATA-binding protein gene regulation apparatus, suggesting that the co-opted full-length bacterial sigma factors are used for specialized instances of gene regulation or serve some other function (Supplementary Information).

The well-described bacterial stringent response based on deployment of multi-domain signalling molecules (guanosine tetraphosphate; ppGpp) called alarmones were identified in one member each of the Diapherotrites and Nanoarchaeota (Fig. 3). These seem to be the result of ancient transfers from bacterial donors of key ppGpp synthetic genes belonging to the RelA/SpoT homologue (RSH) superfamily⁴⁸ (Supplementary Fig. 17 and Supplementary Table 9). Although putative single domain alarmones (synthases and hydrolases) have been found in a number of Euryarchaeota⁴⁸, this is the first report of complete multi-domain archaeal alarmones comprising synthetase, hydrolase and regulatory domains, suggesting that some DPANN Archaea can produce ppGpp in response to the sensation of an intracellular signal. Finally, a bacterial-like lytic murein transglycosylase was found in two members of the Nanoarchaeota (Fig. 3 and Supplementary Fig. 18). This enzyme is ubiquitous in Bacteria and responsible for creating space within the peptidoglycan sacculus for its biosynthesis, recycling and cell division and is tightly regulated because of its potent activity⁴⁹. As Archaea lack peptidoglycan and there is no evidence for peptidoglycan synthesis in the Nanoarchaeota, we speculate that the murein transglycosylase is secreted from the cell and used as a defensive mechanism against bacteria or possibly as a mechanism for facilitating cell-to-cell interaction with bacteria.

Phylogenetic anchoring of metagenomes

A major challenge in metagenomics is determining the phylogenetic origin of anonymous genome fragments, a process called binning or classification⁵⁰. Our ability to classify metagenomic fragments is hampered by the enormous under-sampling of MDM reflected in a highly biased reference genome data set (Supplementary Fig. 1b). To determine whether our set of phylogenetically novel single-cell genomes improves metagenomic binning, we classified 893 publicly available metagenomes against the non-redundant database with and without the 201 SAGs (the single-cell genomes constitute a minimal increase in total database size of 0.7%). Over half (475) of these metagenomes showed new or improved read anchoring (Supplementary Table 10), which accounted for a total of 340 million reads (0.7%). Although this average percentage may seem small, up to 20% anchoring was achieved for some metagenomes, reinforcing the need for phylogenetically directed genomic characterization of microbial diversity. Metagenomes with MDM-SAG-enabled read anchoring of $>2\%$ are shown in Fig. 4, and all other metagenomes are shown in Supplementary Table 11. On average, BLASTX matches of the 340 million reclassified reads increased by approximately 27% amino acid identity, resulting in higher resolution assignments for two-thirds of these reads. Of these, 78% and

22% were newly assigned or re-assigned at the phylum level, respectively (Supplementary Fig. 19 and Supplementary Table 10). The most pronounced improvements were seen in habitats comprising dominant populations belonging to phyla that are well represented in the SAG data set including the Marinimicrobia (SAR406), Aminacnantes (OP8), Cloacimonetes (WWE-1), Parcubacteria (OD1), Atribacteria (OP9) and Microgenomates (OP11) (Fig. 4). Despite these improvements, the majority of reads in the 475 metagenomes could not be classified beyond domain level (up to 80% in some metagenomes) attesting to the continuing need for MDM exploration.

Outlook

Increasing genomic coverage of the microbial world has emerged as a major goal over the past decade and notable international efforts are underway; for example, the Microbial Earth Project, which aims to generate a comprehensive genome catalogue of all archaeal and bacterial type strains (<http://www.microbial-earth.org>), and the Earth Microbiome Project, which uses metagenomics, metatranscriptomics and amplicon sequencing to analyse microbial communities across the globe (<http://www.earthmicrobiome.org>). Although these projects will undoubtedly increase our understanding and appreciation of the microbial world, the phylogenetically targeted approach applied in the GEBA project⁵ and in the present study complements these efforts and facilitates novel discovery. For example, our single-cell genome data set provides an 11% greater coverage of known phylogenetic diversity than currently available genomes according to SSU

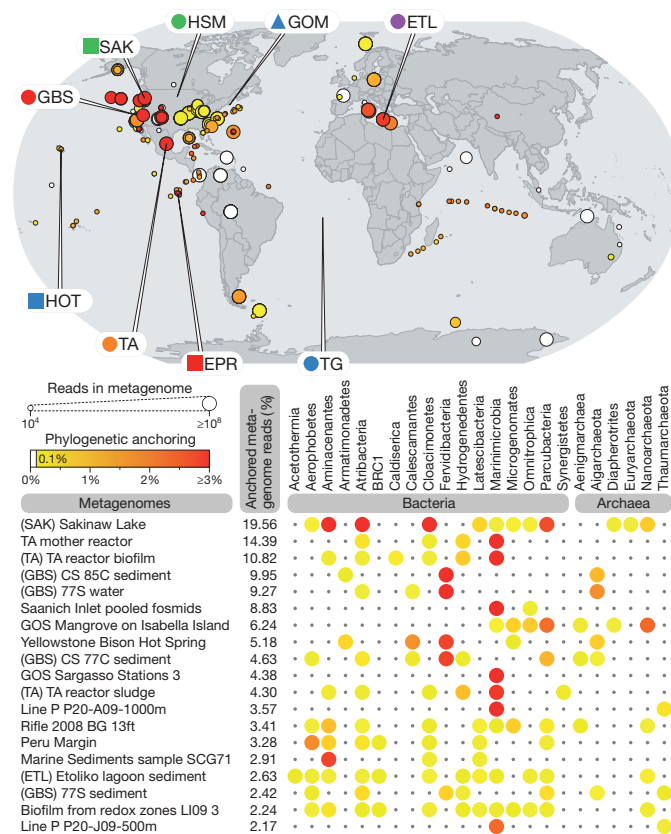


Figure 4 | Phylogenetic anchoring. The geographic location of all 475 metagenome sample sites (circles) and the origin of the MDM samples from which the SAGs were derived. The heatmap below the world map shows the details of 19 metagenomes whose phylogenetic anchoring could be improved for at least 2% of all reads. Phylogenetic anchoring was calculated as the percentage of reads that could be assigned to novel phyla using MEGAN4 results based on BLASTX analysis of all metagenomes against the NCBI non-redundant database before and after addition of MDM data. Statistical testing revealed a significant ($P = 0.00024$) increase in reads that were anchored beyond domain level after the addition of MDM data.

rRNA comparisons (Supplementary Fig. 20a). This represents a 4.5-fold increase in phylogenetic diversity per genome relative to the average phylogenetic diversity of genomes in the public database and a twofold phylogenetic diversity increase per genome afforded by GEBA⁵ (Supplementary Fig. 20a). This increase is also reflected in overall protein novelty with nearly 20,000 new hypothetical protein families in the GEBA-MDM data set, representing an increase of 8.5% compared to the number of genomes sequenced to date (Supplementary Fig. 21). Although the phylogenetic diversity of microbial isolates has increased gradually over time as pure cultures accrue, the phylogenetic diversity of uncultivated microorganisms identified in SSU rRNA surveys has quadrupled since 2007 and currently represents >85% of known microbial diversity (Supplementary Fig. 20b). We estimate that a sequencing effort of at least 16,000 additional genomes from diverse environments is needed to cover 50% of the known phylogenetic diversity based on SSU rRNA profiling (Supplementary Fig. 20a). Single-cell genomics offers a means to inventory this genomic diversity at the organism level directly, bypassing the assembly and binning problems associated with plurality sequencing approaches. Further development of single-cell technologies should overcome known challenges such as fragmented genome recoveries⁸ and will make this technique a more robust tool. As single-cell and other cultivation-independent genomic approaches are used, we anticipate robust improvements to the genomic tree of life that will supercede the single-locus resolution of the SSU rRNA tree. As the genomic tree is filled in, we will witness for the first time a global view of the evolutionary forces that have shaped life on Earth.

METHODS SUMMARY

Nine sites were sampled for single-cell sorting, whole-genome amplification and SSU rRNA screening. A total of 201 phylogenetically targeted SAGs were shotgun sequenced and assembled. Genome completeness was estimated based on universal, single-copy genes. Genome trees were calculated from concatenated alignments of up to 38 universally conserved protein-coding genes in Bacteria and Archaea, and phylogenetic inference was carried out via RAxML, RAxML-Light, and fasttree using 15 taxon configurations. Gene predictions, functional annotation, manual curation and pathway reconstruction were carried out within the Integrated Microbial Genomes (IMG) system (<http://img.jgi.doe.gov>). Phylogenetic anchoring of metagenomic reads was computed using protein blast and the lowest common ancestor approach. Phylogenetic diversity values were calculated from a SSU rRNA maximum likelihood tree. All steps are detailed in the Supplementary Information.

Received 14 January; accepted 4 June 2013.

Published online 14 July 2013.

- Rajendhran, J. & Gunasekaran, P. Microbial phylogeny and diversity: small subunit ribosomal RNA sequence analysis and beyond. *Microbiol. Res.* **166**, 99–110 (2011).
- Gilbert, J. A. & Dupont, C. L. Microbial metagenomics: beyond the genome. *Ann. Rev. Mar. Sci.* **3**, 347–371 (2011).
- Pedros-Alí, C. Marine microbial diversity: can it be determined? *Trends Microbiol.* **14**, 257–263 (2006).
- Hugenholtz, P. & Kyrpides, N. C. A changing of the guard. *Environ. Microbiol.* **11**, 551–553 (2009).
- Wu, D. *et al.* A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* **462**, 1056–1060 (2009).
- Marcy, Y. *et al.* Dissecting biological 'dark matter' with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proc. Natl Acad. Sci. USA* **104**, 11889–11894 (2007).
- Handelsman, J. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **68**, 669–685 (2004).
- Stepanovskas, R. Single cell genomics: an individual look at microbes. *Curr. Opin. Microbiol.* **15**, 613–620 (2012).
- Swan, B. K. *et al.* Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333**, 1296–1300 (2011).
- Konstantinidis, K. T., Ramette, A. & Tiedje, J. M. The bacterial species definition in the genomic era. *Phil. Trans. R. Soc. Lond. B* **361**, 1929–1940 (2006).
- Zaneveld, J. R., Lozupone, C., Gordon, J. I. & Knight, R. Ribosomal RNA diversity predicts genome diversity in gut bacteria and their relatives. *Nucleic Acids Res.* **38**, 3869–3879 (2010).
- Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).
- Szöllosi, G. J., Boussau, B., Abby, S. S., Tannier, E. & Daubin, V. Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations. *Proc. Natl Acad. Sci. USA* **109**, 17513–17518 (2012).
- Ciccarelli, F. D. *et al.* Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**, 1283–1287 (2006).
- Dalevi, D., Hugenholtz, P. & Blackall, L. A multiple-outgroup approach to resolving division-level phylogenetic relationships using 16S rDNA data. *Int. J. Syst. Evol. Microbiol.* **51**, 385–391 (2001).
- Bergsten, J. A review of long-branch attraction. *Cladistics* **21**, 163–193 (2005).
- Gordon, D. A. & Giovannoni, S. J. Detection of stratified microbial populations related to *Chlorobium* and *Fibrobacter* species in the Atlantic and Pacific oceans. *Appl. Environ. Microbiol.* **62**, 1171–1177 (1996).
- Hugenholtz, P., Pitulle, C., Hershberger, K. L. & Pace, N. R. Novel division level bacterial diversity in a Yellowstone hot spring. *J. Bacteriol.* **180**, 366–376 (1998).
- Dojka, M. A., Hugenholtz, P., Haack, S. K. & Pace, N. R. Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. *Appl. Environ. Microbiol.* **64**, 3869–3877 (1998).
- McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610–618 (2012).
- Wagner, M. & Horn, M. The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. *Curr. Opin. Biotechnol.* **17**, 241–249 (2006).
- Gupta, R. S., Bhandari, V. & Naushad, H. S. Molecular signatures for the PVC clade (Planctomycetes, Verrucomicrobia, Chlamydiae, and Lentisphaerae) of bacteria provide insights into their evolutionary relationships. *Front. Microbiol.* **3**, 327 (2012).
- Gupta, R. S. The phylogeny and signature sequences characteristics of Fibrobacteres, Chlorobi, and Bacteroidetes. *Crit. Rev. Microbiol.* **30**, 123–143 (2004).
- Zhang, H. *et al.* Gemmatimonas aurantiaca gen. nov., sp. nov., a gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum Gemmatimonadetes phyl. nov. *Int. J. Syst. Evol. Microbiol.* **53**, 1155–1163 (2003).
- Miroshnichenko, M. L. *et al.* *Caldithrix abyssi* gen. nov., sp. nov., a nitrate-reducing, thermophilic, anaerobic bacterium isolated from a Mid-Atlantic Ridge hydrothermal vent, represents a novel bacterial lineage. *Int. J. Syst. Evol. Microbiol.* **53**, 323–329 (2003).
- Battistuzzi, F. U. & Hedges, S. B. A major clade of prokaryotes with ancient adaptations to life on land. *Mol. Biol. Evol.* **26**, 335–343 (2009).
- Tamaki, H. *et al.* *Armatimonas rosea* gen. nov., sp. nov., a Gram-negative, aerobic, chemoheterotrophic bacterium of a novel bacterial phylum, *Armatimonadetes* phyl. nov., formally called the candidate phylum OP10. *Int. J. Syst. Evol. Microbiol.* **61**, 1442–1447 (2011).
- Sutcliffe, I. C. Cell envelope architecture in the Chloroflexi: a shifting frontline in a phylogenetic turf war. *Environ. Microbiol.* **13**, 279–282 (2011).
- Albertsen, M. *et al.* Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnol.* **6**, 533–538 (2013).
- Chouari, R. *et al.* Novel major bacterial candidate division within a municipal anaerobic sludge digester. *Appl. Environ. Microbiol.* **71**, 2145–2153 (2005).
- Harris, J. K., Kelley, S. T. & Pace, N. R. New perspective on uncultured bacterial phylogenetic division OP11. *Appl. Environ. Microbiol.* **70**, 845–849 (2004).
- Ley, R. E. *et al.* Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Appl. Environ. Microbiol.* **72**, 3685–3695 (2006).
- Wrighton, K. C. *et al.* Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337**, 1661–1665 (2012).
- Baker, B. J. *et al.* Enigmatic, ultrasmall, uncultivated Archaea. *Proc. Natl Acad. Sci. USA* **107**, 8806–8811 (2010).
- Narasimharao, P. *et al.* De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J.* **6**, 81–93 (2012).
- Takai, K. & Horikoshi, K. Genetic diversity of Archaea in deep-sea hydrothermal vent environments. *Genetics* **152**, 1285–1297 (1999).
- Takai, K., Moser, D. P., DeFlaun, M., Onstott, T. C. & Fredrickson, J. K. Archaeal diversity in waters from deep south african gold mines. *Appl. Environ. Microbiol.* **67**, 5750–5760 (2001).
- Guy, L. & Ettema, T. J. G. The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol.* **19**, 580–587 (2011).
- Lake, J. A., Henderson, E., Oakes, M. & Clark, M. W. Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc. Natl Acad. Sci. USA* **81**, 3786–3790 (1984).
- Williams, T. A., Foster, P. G., Nye, T. M. W., Cox, C. J. & Embley, T. M. A congruent phylogenomic signal places eukaryotes within the Archaea. *Proc. R. Soc. B* **279**, 4870–4879 (2012).
- Campbell, J. H. *et al.* UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *Proc. Natl Acad. Sci. USA* **110**, 5540–5545 (2013).
- Johansson, L., Gafvelin, G. & Arnér, E. S. J. Selenocysteine in proteins—properties and biotechnological use. *Biochimica et Biophysica Acta* **1726**, 1–13 (2005).
- Yamamoto, F. *et al.* UGA is read as tryptophan in *Mycoplasma capricolum*. *Proc. Natl Acad. Sci. USA* **82**, 2306–2309 (1985).
- McCutcheon, J. P., McDonald, B. R. & Moran, N. A. Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet.* **5**, e1000565 (2009).
- Zhang, Y., Morar, M. & Ealick, S. E. Structural biology of the purine biosynthetic pathway. *Cell. Mol. Life Sci.* **65**, 3699–3724 (2008).

46. Kyrpides, N. C. & Ouzounis, C. A. Bacterial sigma 70 transcription factor DNA-binding domains in the archaeon *Methanococcus jannaschii*. *J. Mol. Evol.* **45**, 706–707 (1997).
47. Paget, M. S. & Helmann, J. D. The $\sigma 70$ family of sigma factors. *Genome Biol.* **4**, 203 (2003).
48. Atkinson, G. C., Tenson, T. & Hauryliuk, V. The RelA/SpoT homolog (RSH) superfamily: distribution and functional evolution of ppGpp synthetases and hydrolases across the tree of life. *PLoS ONE* **6**, e23479 (2011).
49. Scheurwater, E., Reid, C. W. & Clarke, A. J. Lytic transglycosylases: bacterial space-making autolysins. *Int. J. Biochem. Cell Biol.* **40**, 586–591 (2008).
50. Dröge, J. & McHardy, A. C. Taxonomic binning of metagenome samples generated by next-generation sequencing technologies. *Brief. Bioinform.* **13**, 646–655 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the DOE JGI production sequencing, IMG and GOLD teams for their support; J. Lee and E. Ng for experimental assistance; H.-P. Klenk and D. Gleim for providing a DSMZ inventory database dump and I. Letunić for his knowledge and support to make iTOL work for this project. We are very grateful to B. Schink for invaluable etymological advice. The work conducted by the US Department of Energy Joint Genome Institute is supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231. We also thank the CeBiTec Bioinformatics Resource Facility, which is supported by BMBF grant 031A190. B.P.H. and J.A.D. were supported by the NASA Exobiology grant EXO-NNX11AR78G and NSF OISE 096842 and B.P.H. by a generous contribution from G. Fullmer through the UNLV Foundation. S.M.S. was supported by NSF grants OCE-0452333 and OCE-1136727, and the WHOI's Andrew W. Mellon Fund for Innovative Research; and S.J.H. by the Canadian Foundation for Innovation, the British Columbia Knowledge Development Fund, the National Sciences and Engineering Research Council (NSERC) of Canada and the TULA foundation funded Centre for Microbial Diversity and Evolution (CMDE), and the Canadian Institute for Advanced Research (CIFAR). R.S. was supported by NSF grants DEB-841933, EF-826924,

OCE-1232982, OCE-821374 and OCE-1136488, and the Deep Life I grant by the Alfred P. Sloan Foundation. P.H. was supported by a Discovery Outstanding Researcher Award (DORA) from the Australian Research Council, grant DP120103498.

Author Contributions T.W., C.R. and E.M.R. designed the project, B.K.S., E.A.G., J.A.D., B.P.H., G.T., S.M.S., W.-T.L., S.J.H. and R.S. provided the samples, C.R., T.W. and J.-F.C. performed the experiments, C.R., P.S., A.S., N.N.I., I.J.A., A.D. and S.M. analysed the data, C.R., P.S. and I.J.A. created the figures and tables, and C.R., P.H. and T.W. wrote the manuscript with significant input from A.S., A.D., S.J.H., B.P.H., N.C.K., J.A.E., R.S. and E.M.R.

Author Information Whole-Genome Shotgun projects have been deposited at GenBank under the accession numbers AQP100000000, AQL000000000–AQR00000000, AQA000000000–AQSZ00000000, AQT000000000–AQT000000000, AQL000000000–AQYX00000000, ART000000000, ARW000000000, ASKJ000000000–ASKZ00000000, ASLA000000000–ASLZ00000000, ASMA000000000–ASMZ00000000, ASNA000000000–ASNZ00000000, ASOA000000000–ASOZ00000000, ASPA000000000–ASPH00000000, ASPJ000000000–ASPO00000000, ASWY00000000, ASZK00000000 and ASZL00000000. The annotated single-cell assemblies can be accessed via IMG (<http://img.jgi.doe.gov>). Single-cell genome assemblies are also available at the Microbial Dark Matter project webpage (<http://genome.jgi.doe.gov/MDM>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.W. (twoyke@lbl.gov) and P.H. (p.hughholtz@uq.edu.au).



This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

Structure of class B GPCR corticotropin-releasing factor receptor 1

Kaspar Hollenstein¹, James Kean¹, Andrea Bortolato¹, Robert K. Y. Cheng¹, Andrew S. Doré¹, Ali Jazayeri¹, Robert M. Cooke¹, Malcolm Weir¹ & Fiona H. Marshall¹

Structural analysis of class B G-protein-coupled receptors (GPCRs), cell-surface proteins that respond to peptide hormones, has been restricted to the amino-terminal extracellular domain, thus providing little understanding of the membrane-spanning signal transduction domain. The corticotropin-releasing factor receptor type 1 is a class B receptor which mediates the response to stress and has been considered a drug target for depression and anxiety. Here we report the crystal structure of the transmembrane domain of the human corticotropin-releasing factor receptor type 1 in complex with the small-molecule antagonist CP-376395. The structure provides detailed insight into the architecture of class B receptors. Atomic details of the interactions of the receptor with the non-peptide ligand that binds deep within the receptor are described. This structure provides a model for all class B GPCRs and may aid in the design of new small-molecule drugs for diseases of brain and metabolism.

G-protein-coupled receptors (GPCRs) transmit extracellular signals from hormones, neurotransmitters and metabolites across the plasma membrane and can be classified into four major families based on sequence similarity^{1–3}: A (rhodopsin), B (secretin), C (metabotropic glutamate) and frizzled. Class B, the second largest class, is further divided into the secretin and adhesion subfamilies. The secretin subfamily consists of 15 GPCRs, including receptors for peptides such as secretin, glucagon, glucagon-like peptide (GLP), calcitonin gene-related peptide, parathyroid hormone (PTH) and corticotropin-releasing factor. They are drug targets for the treatment of major diseases including diabetes, osteoporosis, migraine, depression and anxiety. Many class B GPCRs have been clinically validated through the use of injectable peptide drugs derived from their native ligands, such as exenatide, a glucagon-like peptide 1 (GLP-1) receptor agonist for diabetes⁴, and teriparatide, a PTH1 receptor agonist for osteoporosis⁵. However, it has proved extremely difficult to find small-molecule modulators of these receptors.

Since 2000 the structures of more than 18 class A receptors and one member of the frizzled class have been solved, greatly adding to our understanding of the function of GPCRs at a molecular level^{6,7}. In addition, the availability of high-resolution structures in complex with small-molecule ligands has enabled the use of structure-based drug design techniques to be applied to GPCRs for the first time⁸. In contrast to class A receptors, structural information on class B receptors has been limited to the large amino-terminal extracellular domain (ECD)^{9–15}, and no structure of a class B transmembrane domain (TMD), the main target for small-molecule drugs, has been determined to date.

Corticotropin-releasing factor (CRF, also known as corticotropin-releasing hormone, CRH) is a 41 amino acid peptide hormone and a key regulator of the stress response^{16–18}. Through activation of the hypothalamic–pituitary axis and its role as a neurotransmitter, CRF influences a wide range of physiological responses, including appetite control, cardiovascular regulation, glucose metabolism, immune function and behaviour. CRF and the related urocortin peptides (Ucn 1–3, also known as UCN, UCN2 and UCN3) mediate their actions through two mammalian receptor subtypes: CRF₁ and CRF₂. The CRF₁ receptor (CRF₁R, also known as CRHR1), activated by CRF and Ucn1,

is expressed in brain areas including the pituitary, hypothalamus, amygdala and cortex. CRF₁R antagonists have been evaluated for their therapeutic potential in a range of stress-related indications such as anxiety, depression and irritable bowel syndrome¹⁸. A number of structures of the ECD domain of the receptor in complex with CRF have been solved previously⁹. As in other class B GPCRs, the carboxy end of the natural peptide agonist binds primarily to the ECD of the receptor, whereas the amino portion is thought to interact with and engage the TMD, resulting in activation of the receptor¹⁸. In contrast, small-molecule antagonists, such as CP-376395¹⁹, which were discovered by high-throughput screening, seem to act allosterically within the TMD to inhibit the binding and signalling of the peptide-agonist ligands^{20,21}. To advance our understanding of the mode of action of class B GPCRs and to enable structure-based drug design for this important class of receptors, we have determined the crystal structure of the TMD of human CRF₁R.

Structure determination

To obtain a structure of the transmembrane region of CRF₁R, a thermally stabilized receptor (StaR) was generated that preferentially adopts the inactive conformation using an approach described previously^{22–24}. The StaR contains twelve amino acid substitutions, none of which are located in or adjacent to the antagonist-binding site. The crystallized construct lacks the ECD and T4 lysozyme (T4L) has been inserted into intracellular loop (ICL) 2 (Fig. 1). The carboxy terminus was truncated to increase expression levels removing the putative helix 8. This construct shows similar affinity for the 2-aryloxy-4-alkylaminopyridine CP-376395 to the wild-type CRF₁R (Supplementary Table 1). The structure was solved by molecular replacement to 3.0 Å with three receptor molecules (A–C) per asymmetric unit (Supplementary Figs 1–4). The following discussions refer to molecule C, where there was continuous density for amino acids 117–367.

Overall architecture

The core fold of the receptor features seven transmembrane helices (TM1–TM7) in a generally similar arrangement to those observed in previously determined GPCR structures (Fig. 2). The connecting

¹Heptares Therapeutics Ltd, BioPark, Broadwater Road, Welwyn Garden City, AL7 3AX, UK.

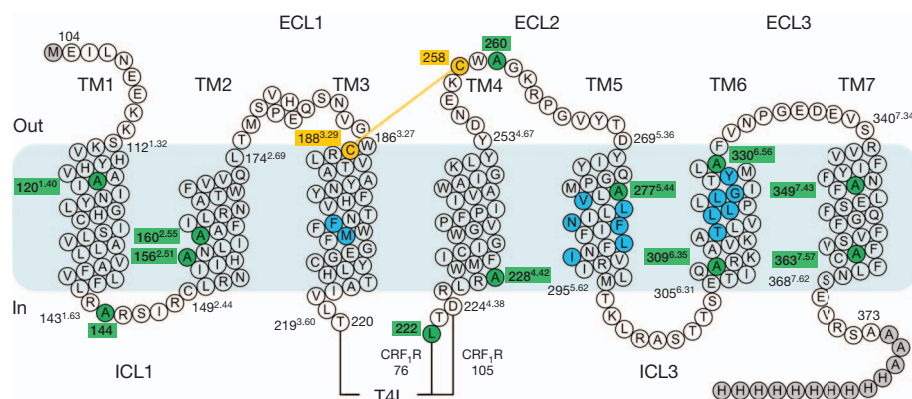


Figure 1 | Schematic representation of the CRF₁R structure. For crystallization, both termini of human CRF₁R were truncated resulting in constructs of the TMD (residues 104–373) lacking the ECD involved in peptide binding. Thermostabilizing mutations and residues involved in small-molecule antagonist binding are highlighted in green and blue, respectively. Other

modified sequence is coloured in grey. The disulphide bond between Cys 258 in ECL2 and Cys 188^{3.29} at the top of TM3 is indicated by a yellow line. The sites of the T4L insertions into ICL2 are indicated (see Methods for details on constructs CRF₁R 76 and CRF₁R 105). The first and last residues of each TM helix are labelled (Wootten numbering in superscript²⁶).

loops lack secondary structure apart from extracellular loop (ECL) 1, which contains a short α -helix parallel to the membrane. A helical structure of ECL1 has been proposed in an NMR study²⁵ on the PTH1 receptor and might therefore be an architectural feature conserved in class B GPCRs. ECL2 is anchored to TM3 by a disulphide bond between Cys 188^{3.29} (Wootten numbering²⁶ in superscript) and Cys 258, two fully conserved residues in class B receptors. A similar disulphide bond, also found in most class A GPCRs²⁷, is reported to be important for agonist-induced receptor activation in the class B GPCR GLP-1 receptor²⁸.

Comparison with class A GPCRs

Comparison of the CRF₁R structure with previously solved GPCRs provides insight into the architectural differences between class A and

class B receptors. Unlike class A GPCRs, the CRF₁R TMD adopts a pronounced V-shape (Fig. 2b), presenting a large cavity, presumed to be the peptide-binding site, to the extracellular side (Fig. 2d). We illustrate the differences using the dopamine D₃ receptor (D₃R)²⁹ as a representative of class A GPCRs, because its overall fold is determined to be closest to that of CRF₁R (Supplementary Table 2). Superposing the two receptors reveals striking differences in the arrangement of TM7 and smaller, but substantial, differences in TM6 (Fig. 3 and Supplementary Table 2). In both receptors TM7 assumes a similar, sharply kinked structure. In CRF₁R, however, its extracellular end is located approximately 10 Å further away from the axis of the helical bundle, a consequence of a $\sim 25^\circ$ rotation of TM7 relative to D₃R around a pivot near Gly 356^{7.50} (Supplementary Fig. 5 and Supplementary Table 3). Since the carboxy terminus has been truncated to remove helix 8 it is

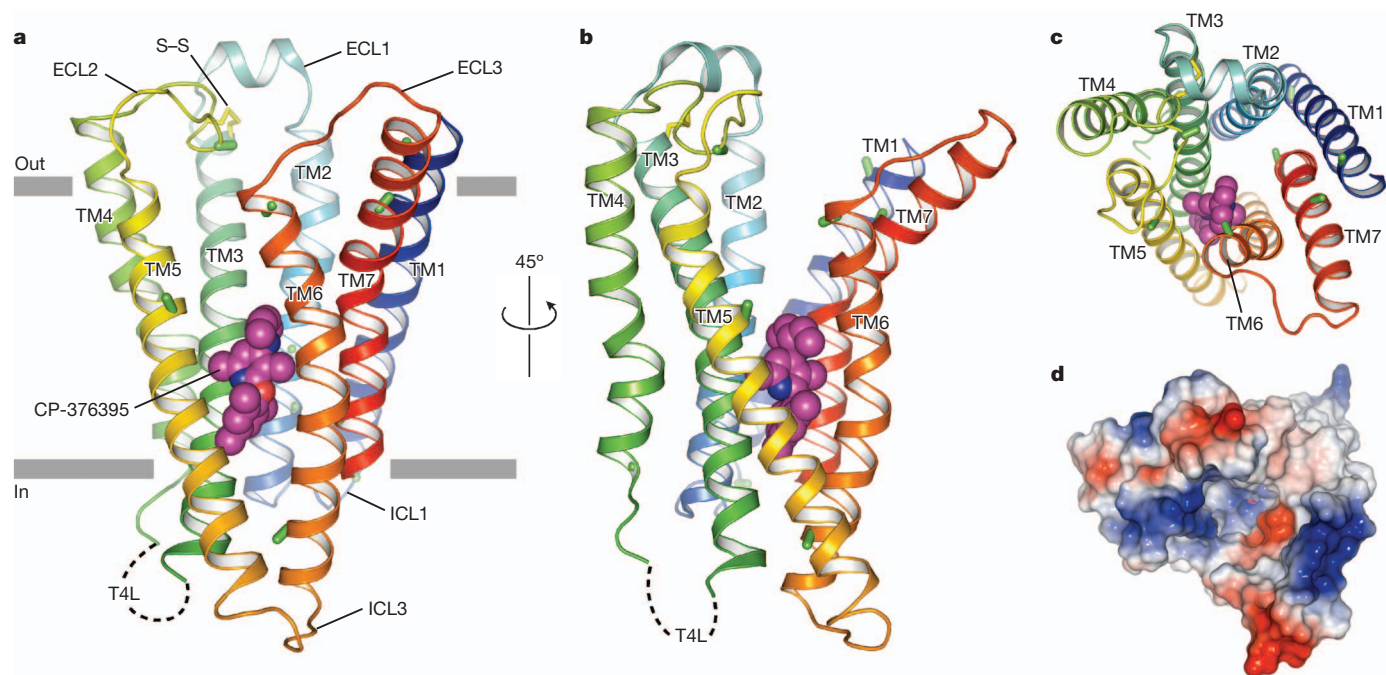


Figure 2 | Overall structure. a, b, Ribbon representation of CRF₁R in rainbow colouration (amino terminus, blue; carboxy terminus, red) viewed from the membrane from two angles. The antagonist CP-376395 is depicted in space fill representation with carbon, nitrogen and oxygen atoms coloured magenta, blue and red, respectively. The disulphide bond linking ECL1 and TM3 is shown as

yellow sticks. The thermostabilizing mutations are rendered as green sticks. The position of the T4L insertion is indicated. c, The receptor is viewed from the extracellular side. d, Same view as in c, but in a surface representation coloured by electrostatic potential.

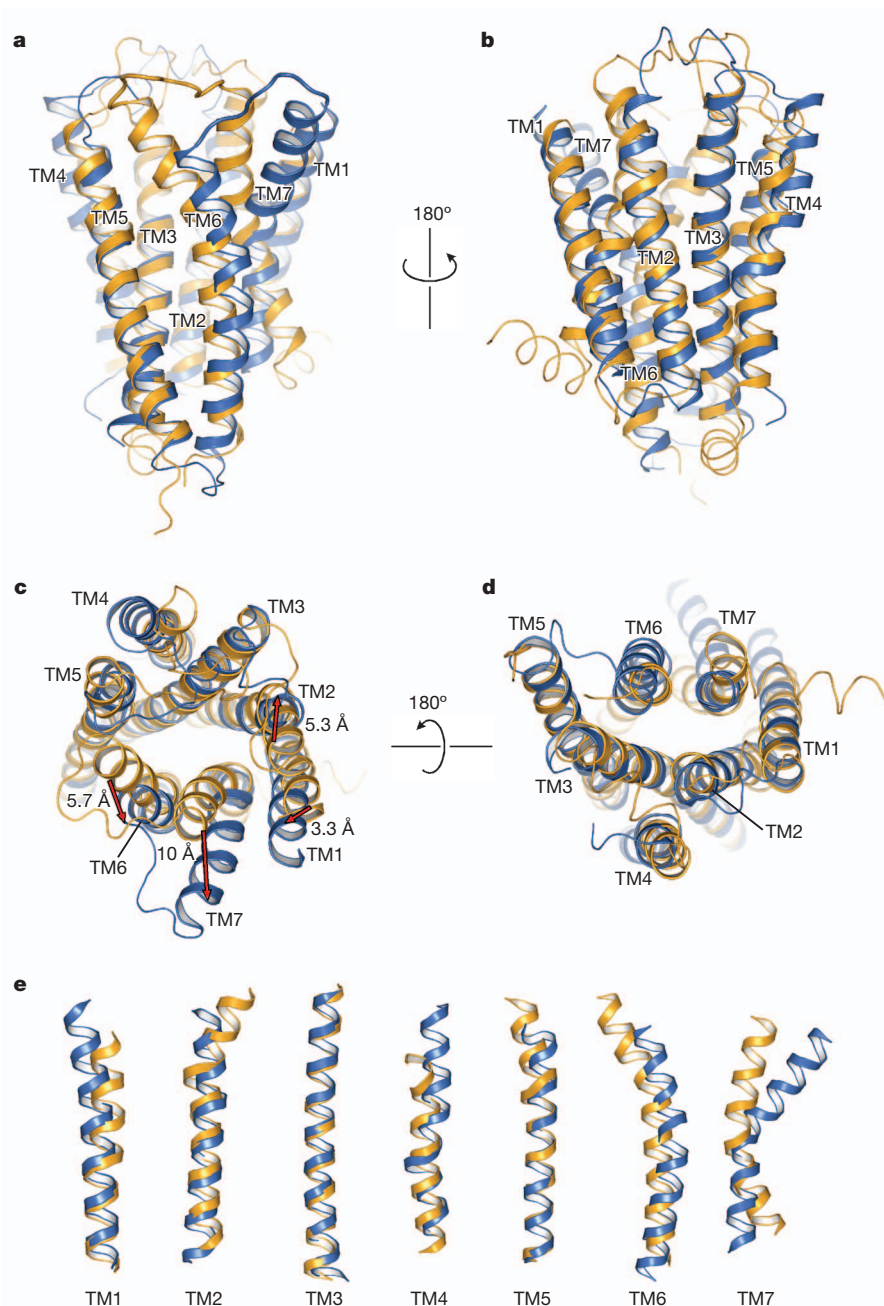


Figure 3 | Comparison of the antagonist-bound structures of CRF₁R and D₃R. a–e, The superimposed structures of CRF₁R and D₃R (PDB ID 3PBL) are shown as blue and yellow ribbons, respectively. The receptors are viewed from two different angles from within the membrane (a, b), from the extracellular

side (c), and intracellular side (d). Arrows in c highlight large differences in the helical trajectories of the two receptors. e, Individual TM helices are shown after superposition of the two receptors as in a–d. The superposition of D₃R onto CRF₁R was done as described in Methods.

possible that this has affected the position of TM7 observed in the crystal structure; however, this truncation has no effect on the ligand-binding properties of the receptor (Supplementary Table 1). TM6 in CRF₁R is also shifted away from TM5 and the axis of the helical bundle. In addition, its extracellular end is shorter by two helical turns, which limits the interactions between TM5 and TM6 to their cytoplasmic regions.

The slightly bent extracellular portion of TM1 packs against TM7 and hence follows a similar trajectory away from the centre of the receptor. The highly conserved Ser 130^{1.50} stabilizes the kink in TM7 through hydrogen bonds to the backbone at Phe 357^{7.51} and Ser 353^{7.47} (Fig. 4a and Supplementary Fig. 6). A similar interaction is found in class A receptors where an asparagine at position 1.50

(Ballesteros–Weinstein numbering³⁰), located one helical turn down relative to Ser 130^{1.50} in CRF₁R, binds to the backbone of TM7. Whereas in most class A receptors the extracellular tip of TM2 leans towards TM1, in CRF₁R TM2 is straighter, thereby contributing to the open nature of the receptor. Because the structure solved here lacks the ECD it does not yet provide an understanding of the relative positions of the ECD and TMD, nor the mechanism of activation following ligand binding. A two-step process has been proposed for class B receptor activation. First the amino-terminal ECD of the receptor binds the carboxy terminus of the peptide ligand; this results in a rearrangement of the receptor domains with the insertion of the peptide agonist amino terminus into the transmembrane helix bundle, triggering conformational changes within the TMD to the active G-protein-binding state¹⁸. The

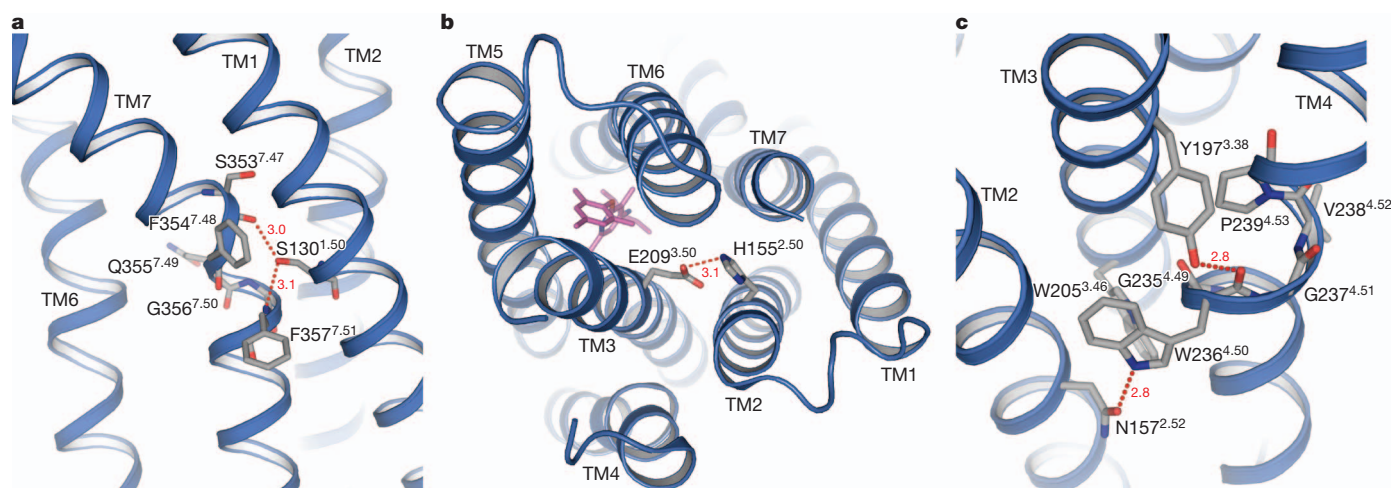


Figure 4 | Conserved sequence motifs. a–c, CRF₁R is shown in blue ribbons with interacting residues shown as sticks with carbon, nitrogen and oxygen atoms coloured grey, blue and red, respectively. Interacting residues and TM

helices are labelled. Hydrogen bonds are depicted as red dotted lines and their lengths in Å are indicated. CP-376395 is represented in sticks and coloured as in Fig. 2a.

open structure of the TMD observed in the structure may relate to the way the large peptide agonist in complex with the ECD interacts with the TMD.

Class A and class B GPCRs signal through the same effector proteins, which interact with the cytoplasmic side of the receptors. Comparison of CRF₁R with D₃R reveals that, unlike their extracellular portions, their cytoplasmic regions superimpose well (Fig. 3d and Supplementary Table 3). Specifically, the cytoplasmic halves of TM3 and TM5, which have been found to interact with G α_s in the structure of the β_2 -adrenergic receptor–G α_s complex³¹, adopt similar conformations in CRF₁R and D₃R. Despite the lack of a native ICL2 due to the T4L insertion, the relative arrangement of the intracellular half of TM4 with respect to that of TM3 is similar to that in D₃R. This would bring ICL2 into a similar position to class A receptors, thus allowing receptor–G protein interaction. Even though TM6 in CRF₁R is shifted outward, its cytoplasmic end points towards TM3 in a similar fashion to the structures of D₃R and other class A receptors solved in an inactive state²⁷ (Supplementary Fig. 7), indicating that the observed conformation of the CRF₁R also represents the inactive state.

Interactions in conserved motifs

In class A GPCRs, a conserved salt bridge connects TM6 to TM3 in the inactive conformation. Class B receptors lack the sequence motifs for this ‘ionic lock’. Instead, biochemical data suggest interaction of His 155^{2.50} and Glu 209^{3.50} to be involved in activation^{32–35}. In CRF₁R, these side chains are within hydrogen-bonding distance, forming a potentially important energetic interaction (Fig. 4b and Supplementary Fig. 6). In class B GPCRs, a conserved sequence motif, GWGxP, is present in TM4. The structure reveals a network of interactions around Trp 236^{4.50} of this motif (Fig. 4c and Supplementary Fig. 6). TM4 unwinds at Gly 235^{4.49}, resulting in Trp 236^{4.50} protruding towards TM2 and TM3. This side chain hydrogen bonds to Asn 157^{2.52} and forms an edge-to-face interaction with Trp 205^{3.46}. The TM4–TM3 interaction is strengthened by a hydrogen bond from the side chain of Tyr 197^{3.38} to the carbonyl of Trp 236^{4.50} and hydrophobic contacts of Gly 235^{4.49} with Trp 205^{3.46} and Pro 239^{4.53} with Tyr 197^{3.38}. These conserved interactions seem to have an important structural role. Whether they are also involved in receptor function is not immediately clear from the structure.

Small-molecule antagonist-binding site

Strong electron density was found for CP-376395, unexpectedly in a pocket located in the cytoplasmic half of the receptor (Fig. 5). The position is remarkable, as it is approximately 18 Å away from the centre of the large cavity presented to the extracellular side, the

putative peptide agonist-binding site, and 13–23 Å away from the centres of mass of the small-molecule ligands in the structures of class A GPCRs (Fig. 5d). In an orientation perpendicular to the plane of the membrane, CP-376395 binds in a predominantly hydrophobic pocket defined by residues of TM3, TM5 and TM6. The side chain of highly conserved Asn 283^{5.50} forms an essential hydrogen bond with the pyridine nitrogen, such that substitution with alanine results in a complete loss of ligand binding (Supplementary Table 4). The aryloxy moiety binds in a hydrophobic pocket formed by Phe 284^{5.51}, Leu 287^{5.54}, Ile 290^{5.57}, Tyr 316^{6.42}, Leu 319^{6.45} and Leu 320^{6.46}. The exocyclic alkylamino group interacts with Gly 324^{6.50} and the side chains of Phe 203^{3.44}, Leu 280^{5.47}, Leu 323^{6.49} and Tyr 327^{6.53}. To validate the interactions, the amino acids in the antagonist-binding site were substituted with alanine (Supplementary Table 4). Apart from Leu 319^{6.45}, all of these mutations resulted in a significant reduction of antagonist binding with negligible binding detected for Phe203^{3.44} Ala, Leu280^{5.47} Ala and Asn283^{5.50} Ala.

At the bottom of the putative orthosteric site, Arg 165^{2.60}, His 199^{3.40}, Met 276^{5.43} and Gln 355^{7.49} form a layer of side chains just above the antagonist-binding site (Fig. 5e). Mutation of His 199^{3.40} and Met 276^{5.43} has been shown to impair non-peptide antagonist binding to CRF₁R³⁶. Although not interacting with CP-376395 directly, His 199^{3.40} forms a hydrogen bond with Tyr 327^{6.53} and Met 276^{5.43} packs against Phe 203^{3.44}, potentially supporting the positioning of these two aromatic residues. Substitution of highly conserved Gln 355^{7.49} with alanine did not affect CP-376395 binding (Supplementary Table 4). However, this residue may make an energetic contribution to receptor activation as suggested by mutagenesis studies^{37–40}.

The sequence identity in the helical bundle between CRF₁R and CRF₂R (also known as CRHR2) is very high; however, non-peptide antagonists such as CP-376395 are highly selective for CRF₁R. All the residues found to interact directly with the ligand are conserved between the two receptors. Among the second shell of amino acids around the small molecule (within 5 Å) only two residues differ between CRF₁R and CRF₂R. His 199^{3.40} and Met 276^{5.43} in CRF₁R correspond to Val 195^{3.40} and Ile 272^{5.43} of CRF₂R. In CRF₁R, mutation of these residues to the corresponding amino acids in CRF₂R has been shown to reduce the binding affinity of non-peptide antagonists while having no effect on peptide ligand binding³⁶. Although these residues may have an indirect effect on ligand binding they could also act to control the accessibility to the small-molecule antagonist-binding site through affecting the side-chain orientation of the adjacent Phe 203^{3.44} that interacts directly with CP-376395.

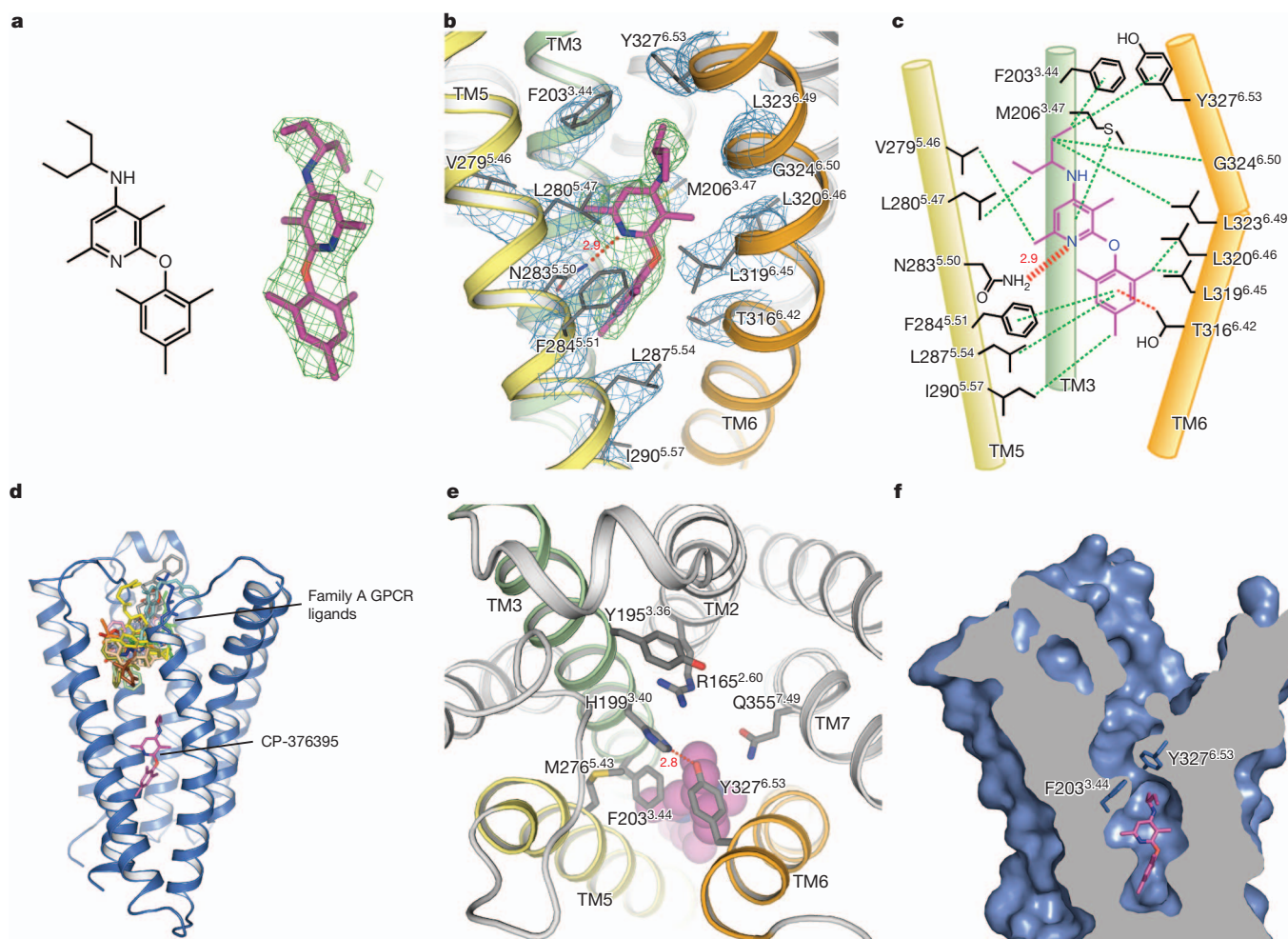


Figure 5 | Antagonist-binding site. **a**, CP-376395 shown as skeletal formula and as sticks coloured as in Fig. 2a. $F_o - F_c$ OMIT density calculated before CP-376395 was first included in the model displayed within 2 \AA , contoured at 2.5σ . **b**, View from the membrane with $2F_o - F_c$ density within 2 \AA of the binding site residues, contoured at 1.5σ . OMIT density for CP-376395 as in **a**. The hydrogen bond to the pyridine nitrogen of CP-376395 is

depicted as a red dotted line and its length in \AA is indicated. **c**, Schematic of the binding site with hydrogen bonds and hydrophobic interactions as red and green dotted lines, respectively. **d**, The location of CP-376395 is compared to that of selected class A receptor ligands. **e**, View from the extracellular side onto the bottom of the putative orthosteric site. **f**, Cross-section of the solvent-accessible surface.

In the β_2 -adrenergic receptor, the cytoplasmic half of TM6 was shown to have a key role in receptor activation by moving away from the core of the TM bundle³¹. CP-376395 makes extensive contacts with TM3, TM5 and TM6. Specifically, it binds to TM6 on both sides of the kink induced by Pro 321^{6.47} and Gly 324^{6.50}. It has been proposed that small-molecule CRF₁R antagonists act allosterically by maintaining the receptor in an inactive conformation²⁰. The position of CP-376395 in the structure indicates that this is indeed an allosteric site distinct from the orthosteric peptide-binding pocket. It is possible that CP-376395 keeps the receptor in an inactive conformation by tethering the cytoplasmic half of TM6 to TM3 and TM5. This proposition is supported by the fact that CP-376395 acts as an inverse agonist reducing basal signalling (Supplementary Fig. 8).

Access to the antagonist-binding site from the extracellular side is restricted to a small channel by the side chains of Phe 203^{3.44} and Tyr 327^{6.53} (Fig. 5f). For CP-376395 to reach the antagonist-binding site from the extracellular side it may require rearrangement of residues at the top of the site through side chain rotamer changes and/or shifts of portions of TM3, TM5 or TM6, which would enlarge the opening towards the putative orthosteric site. Alternatively, the antagonist-binding site, separated from the membrane by a single layer of side chains, could open laterally through rearrangements in

TM5 and TM6, allowing diffusion of the hydrophobic ligand from within membrane.

Conclusions

The structure of the TMD of the CRF₁ receptor provides a structural framework for class B GPCRs and should serve as an accurate template for homology modelling of related receptors. It also reveals an unanticipated antagonist-binding site, opening new avenues for structure-based small-molecule drug discovery. To further our understanding of the mode of action of class B GPCRs, structures of the full-length receptor in the agonist-bound state are now required.

METHODS SUMMARY

Conformationally thermostabilized human CRF₁R lacking the amino-terminal extracellular domain and with T4 lysozyme fused into intracellular loop 2 was expressed in baculovirus-infected insect cells and purified in *n*-decyl- β -D-maltopyranoside using a histidine tag. Crystals were grown in lipidic cubic phase at 22.5°C using a monoolein/cholesterol mixture. Diffraction data from 35 crystals, collected at Diamond Light Source beamline I24, was used to solve the structure by molecular replacement using the structures of T4 lysozyme and the human dopamine D₃ receptor as initial search models. The structure was refined to 3.0 \AA with good refinement statistics ($R_{\text{work}}/R_{\text{free}} = 0.241/0.265$). For details of experimental procedures see Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 4 March; accepted 7 June 2013.

Published online 17 July 2013.

1. Gether, U. Uncovering molecular mechanisms involved in activation of G protein-coupled receptors. *Endocr. Rev.* **21**, 90–113 (2000).
2. Lagerström, M. C. & Schiöth, H. B. Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nature Rev. Drug Discov.* **7**, 339–357 (2008).
3. Venkatakrishnan, A. J. *et al.* Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185–194 (2013).
4. Bhavsar, S., Mudaliar, S. & Cherrington, A. Evolution of exenatide as a diabetes therapeutic. *Curr. Diabetes Rev.* **9**, 161–193 (2013).
5. Berg, C., Neumeyer, K. & Kirkpatrick, P. Teriparatide. *Nature Rev. Drug Discov.* **2**, 257–258 (2003).
6. Rosenbaum, D. M., Rasmussen, S. G. F. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356–363 (2009).
7. Wang, C. *et al.* Structure of the human smoothened receptor bound to an antitumour agent. *Nature* **497**, 338–343 (2013).
8. Congreve, M., Langmead, C. & Marshall, F. H. The use of GPCR structures in drug design. *Adv. Pharmacol.* **62**, 1–36 (2011).
9. Pioszak, A. A., Parker, N. R., Suino-Powell, K. & Xu, H. E. Molecular recognition of corticotropin-releasing factor by its G-protein-coupled receptor CRFR1. *J. Biol. Chem.* **283**, 32900–32912 (2008).
10. Grace, C. R. R. *et al.* NMR structure and peptide hormone binding site of the first extracellular domain of a type B1 G protein-coupled receptor. *Proc. Natl Acad. Sci. USA* **101**, 12836–12841 (2004).
11. Runge, S., Thøgersen, H., Madsen, K., Lau, J. & Rudolph, R. Crystal structure of the ligand-bound glucagon-like peptide-1 receptor extracellular domain. *J. Biol. Chem.* **283**, 11340–11347 (2008).
12. Pioszak, A. A. & Xu, H. E. Molecular recognition of parathyroid hormone by its G protein-coupled receptor. *Proc. Natl Acad. Sci. USA* **105**, 5034–5039 (2008).
13. ter Haar, E. *et al.* Crystal structure of the ectodomain complex of the CGRP receptor, a class-B GPCR, reveals the site of drug antagonism. *Structure* **18**, 1083–1093 (2010).
14. Kumar, S., Pioszak, A., Zhang, C., Swaminathan, K. & Xu, H. E. Crystal structure of the PAC1R extracellular domain unifies a consensus fold for hormone recognition by class B G-protein coupled receptors. *PLoS ONE* **6**, e19682 (2011).
15. Koth, C. M. *et al.* Molecular basis for negative regulation of the glucagon receptor. *Proc. Natl Acad. Sci. USA* **109**, 14393–14398 (2012).
16. Perrin, M. H. & Vale, W. W. Corticotropin releasing factor receptors and their ligand family. *Ann. NY Acad. Sci.* **885**, 312–328 (1999).
17. Bale, T. L. & Vale, W. W. CRF and CRF receptors: role in stress responsivity and other behaviors. *Annu. Rev. Pharmacol. Toxicol.* **44**, 525–557 (2004).
18. Hemley, C. F., McCluskey, A. & Keller, P. A. Corticotropin releasing hormone—a GPCR drug target. *Curr. Drug Targets* **8**, 105–115 (2007).
19. Chen, Y. L. *et al.* 2-aryloxy-4-alkylaminopyridines: discovery of novel corticotropin-releasing factor 1 antagonists. *J. Med. Chem.* **51**, 1385–1392 (2008).
20. Hoare, S. R. *et al.* Allosteric ligands for the corticotropin releasing factor type 1 receptor modulate conformational states involved in receptor activation. *Mol. Pharmacol.* **73**, 1371–1380 (2008).
21. Zorrilla, E. P. & Koob, G. F. Progress in corticotropin-releasing factor-1 antagonist development. *Drug Discov. Today* **15**, 371–383 (2010).
22. Serrano-Vega, M. J., Magnani, F., Shibata, Y. & Tate, C. G. Conformational thermostabilization of the β 1-adrenergic receptor in a detergent-resistant form. *Proc. Natl Acad. Sci. USA* **105**, 877–882 (2008).
23. Shibata, Y. *et al.* Thermostabilization of the neurotensin receptor NTS1. *J. Mol. Biol.* **390**, 262–277 (2009).
24. Lebon, G., Bennett, K., Jazayeri, A. & Tate, C. G. Thermostabilisation of an agonist-bound conformation of the human adenosine A_{2A} receptor. *J. Mol. Biol.* **409**, 298–310 (2011).
25. Piserchio, A., Bisello, A., Rosenblatt, M., Chorev, M. & Mierke, D. F. Characterization of parathyroid hormone/receptor interactions: structure of the first extracellular loop. *Biochemistry* **39**, 8153–8160 (2000).
26. Wootten, D., Simms, J., Miller, L. J., Christopoulos, A. & Sexton, P. M. Polar transmembrane interactions drive formation of ligand-specific and signal pathway-biased family B G protein-coupled receptor conformations. *Proc. Natl Acad. Sci. USA* **110**, 5211–5216 (2013).
27. Katritch, V., Cherezov, V. & Stevens, R. C. Diversity and modularity of G protein-coupled receptor structures. *Trends Pharmacol. Sci.* **33**, 17–27 (2012).
28. Mann, R. J., Al-Sabah, S., De Maturana, R. L., Sinfield, J. K. & Donnelly, D. Functional coupling of Cys-226 and Cys-296 in the glucagon-like peptide-1 (GLP-1) receptor indicates a disulfide bond that is close to the activation pocket. *Peptides* **31**, 2289–2293 (2010).
29. Chien, E. Y. T. *et al.* Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* **330**, 1091–1095 (2010).
30. Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* **25**, 366–428 (1995).
31. Rasmussen, S. G. F. *et al.* Crystal structure of the β 2 adrenergic receptor–Gs protein complex. *Nature* **477**, 549–555 (2011).
32. Schipani, E., Kruse, K. & Jüppner, H. A constitutively active mutant PTH-PTHrP receptor in Jansen-type metaphyseal chondrodysplasia. *Science* **268**, 98–100 (1995).
33. Heller, R. S., Kieffer, T. J. & Habener, J. F. Point mutations in the first and third intracellular loops of the glucagon-like peptide-1 receptor alter intracellular signaling. *Biochem. Biophys. Res. Commun.* **223**, 624–632 (1996).
34. Hjorth, S. A., Orskov, C. & Schwartz, T. W. Constitutive activity of glucagon receptor mutants. *Mol. Endocrinol.* **12**, 78–86 (1998).
35. Vohra, S. *et al.* Similarity between class A and class B G-protein-coupled receptors exemplified through calcitonin gene-related peptide receptor modelling and mutagenesis studies. *J. R. Soc. Interface* **10**, 20120846 (2013).
36. Hoare, S. R. J. *et al.* Single amino acid residue determinants of non-peptide antagonist binding to the corticotropin-releasing factor1 (CRF1) receptor. *Biochem. Pharmacol.* **72**, 244–255 (2006).
37. Gardella, T. J., Luck, M. D., Fan, M. H. & Lee, C. Transmembrane residues of the parathyroid hormone (PTH)/PTH-related peptide receptor that specifically affect binding and signaling by agonist ligands. *J. Biol. Chem.* **271**, 12820–12825 (1996).
38. Conner, A. C. *et al.* A key role for transmembrane prolines in calcitonin receptor-like receptor agonist binding and signalling: implications for family B G-protein-coupled receptors. *Mol. Pharmacol.* **67**, 20–31 (2005).
39. Chugunov, A. O. *et al.* Evidence that interaction between conserved residues in transmembrane helices 2, 3, and 7 are crucial for human VPAC1 receptor activation. *Mol. Pharmacol.* **78**, 394–401 (2010).
40. Ganguli, S. C. *et al.* Protean effects of a natural peptide agonist of the G protein-coupled secretin receptor demonstrated by receptor mutagenesis. *J. Pharmacol. Exp. Ther.* **286**, 593–598 (1998).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank G. Evans, R. Owen and D. Axford for their support. We are grateful to A. Leslie, R. Read and A. McCoy for advice on data collection and structure determination. We thank R. Henderson, A. Leslie, C. Tate as well as F. Blaney, B. Tehan, R. Miller, F. Deflorian and other colleagues for suggestions and comments.

Author Contributions J.K. carried out the conformational thermostabilization and pharmacological characterization of the constructs and determined the stability of the StaR in a panel of reagents/additives to aid purification and crystallization. A.J. constructed the T4 lysozyme fusion proteins and identified the usefulness of the fusion into intracellular loop 2. K.H. designed and characterized truncation constructs, designed the final T4 lysozyme fusion construct, established procedures for, and carried out expression and purification, established the platform/protocols for and carried out lipidic cubic phase (LCP) crystallization, collected and processed X-ray diffraction data, solved and refined the structure. R.K.Y.C. performed expression and purification, optimized purification and grew crystals in LCP for data collection of the final construct, collected X-ray diffraction data and solved and refined the structure. A.S.D. was involved in construct design, established the platform/protocols for, and carried out LCP crystallization, collected and processed X-ray diffraction data and solved and refined the structure. Computational analysis of the structure and modelling was carried out by A.B. Project management was carried out by A.J., R.M.C., M.W. and F.H.M. The manuscript was prepared by K.H., R.K.Y.C., A.B., J.K., A.J. and F.H.M.

Author Information Co-ordinates and structure factors have been deposited in the Protein Data Bank under the accession code 4K5Y. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.H.M. (fiona.marshall@heptares.com).

METHODS

StaR generation. Full-length human CRF₁R with intact ICL2 (no T4L insertion) was used as background for the generation of the conformationally thermostabilized receptor using a mutagenesis approach described earlier⁴¹. Mutants were analysed for thermostability in the presence of the radioligand [³H]CP-376395 (Supplementary Fig. 9a and Supplementary Table 5). The CRF₁R StaR is the full-length receptor with 12 thermostabilizing mutations (see Fig. 1).

Cell culture. HEK293T cells were cultured in DMEM supplemented with 10% (v/v) fetal bovine serum (FBS). Cells were transfected using GeneJuice (Merck Millipore) according to manufacturer's instructions and collected after 48 h.

Thermostability measurement. Transiently transfected HEK293T cells were incubated in 50 mM Tris-HCl pH 7.5, 150 mM NaCl with 30 nM [³H]CP-376395 and 120 nM cold CP-376395 (Tocris) for 18 h at room temperature. All subsequent steps were performed at 4 °C. Cells were solubilized in 1% (w/v) *n*-dodecyl- β -D-maltopyranoside (DDM) for 1 h and crude lysates were cleared by centrifugation at 16,000g for 15 min. Thermostability of the receptor was measured by incubation at varying temperatures for 30 min followed by separation of unbound radioligand by gel filtration. Levels of ligand-bound receptor were determined using a liquid scintillation counter. Thermal stability (T_m) is defined as the temperature at which 50% ligand binding is retained.

Radioligand binding. For saturation binding experiments, membranes were prepared from HEK293T cells transiently expressing wild-type CRF₁R, CRF₁R StaR, CRF₁R StaR with T4L fusion, or CRF₁R 105 as described⁴¹ and incubated in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% (w/v) polyethylenimine (PEI) with [³H]CP-376395 (0–60 nM) in the presence or absence of 30 μ M cold CP-376395. Final dimethyl sulfoxide (DMSO) concentration in each reaction was 5% (v/v). Membranes were incubated for 18 h at room temperature before rapid filtration through 96-well GF/C UniFilter plates pre-soaked in 0.3% (w/v) PEI, followed by washing with PBS containing 0.15% (w/v) CHAPS. Plates were dried, 50 μ l Ultima Gold-F scintillation fluid added per well and bound ligand measured using a Packard Microbeta counter. To obtain K_d , data were analysed using a global fitted one-site binding hyperbola in GraphPad Prism v5. For solubilized whole-cell ligand-binding experiments, HEK293T cells transiently expressing enhanced green fluorescent protein (EGFP)-tagged wild-type CRF₁R or single point mutants were treated as described for thermostability experiments, without heating. Specific binding was determined by subtracting untransfected controls. Expression of each construct was quantified by EGFP fluorescence of whole cells measured at λ_{ex} and λ_{em} of 488 and 520 nm, respectively.

Measurement of cAMP. HEK293T cells were transfected with plasmids bearing wild-type CRF₁R. In parallel mock transfections were set up using empty plasmid. 24 h post-transfection, cells were collected and used to seed white poly-D-lysine coated 96-well plates at a density of 25,000 cells per well. Cells were incubated for a further 24 h at 37 °C. Prior to measuring cAMP levels, CP-376395 was added to a final concentration of 1 μ M. Equivalent DMSO concentration was added to control samples. Following a 1.5 h incubation at 37 °C basal cAMP levels were measured using Cisbio International Cyclic AMP dynamic 2 kit according to the manufacturer's instructions (Supplementary Fig. 8).

Truncation and T4L lysozyme fusion constructs. A panel of amino- and carboxy-terminal truncation variants of CRF₁R was designed based on secondary structure prediction⁴² and hydropathy plots⁴³. Truncated receptors were expressed in HEK293T cells as carboxy-terminal fusions with EGFP. Receptors were solubilized in 50 mM Tris-HCl pH 8.0, 150 mM NaCl and 2% (w/v) *n*-decyl- β -D-maltopyranoside (DM) and their expression levels and stability was assayed by fluorescence-detection size-exclusion chromatography (FSEC) as described⁴⁴ (Supplementary Fig. 9b). The most suitable construct emerging from this screen comprised residues 104–373. In parallel, a panel of T4L insertions into the predicted locations of ICL2 or ICL3 was analysed in a similar fashion, identifying the insertion between residues 220 and 222 in ICL2 as the most promising fusion (construct CRF₁R 76) based on FSEC analysis. Crystals obtained from this construct, however, did not allow structure determination, likely due to the presence of translational non-crystallographic symmetry (see below). To obtain the P22₁2₁ crystal form, which enabled structure determination, residues 222 and 223 were later deleted, yielding the final construct CRF₁R 105 (Fig. 1).

Expression and purification. CRF₁R carrying a carboxy-terminal deca-histidine tag was expressed in *Trichoplusia ni* (High Five) cells in EX-CELL 405 medium supplemented with 10% (v/v) FBS, 1% (v/v) Chemically Defined Lipid Concentrate (GIBCO) and 1% (v/v) penicillin/streptomycin. Cells were infected at a density of 2×10^6 cells per ml with baculovirus at an approximate multiplicity of infection of 1. Cultures were grown at 27 °C and collected 72 h post infection. All subsequent purification steps were carried out at 4 °C unless indicated differently. To prepare membranes, cells were resuspended in 50 mM Tris-HCl pH 8.0, 500 mM NaCl and incubated with 0.3 μ M CP-376395 for 1 h. Cells were disrupted by ultra-sonication and cell debris was removed by centrifugation at 10,000g.

Membranes were collected by ultracentrifugation at 140,000g, resuspended and stored at –80 °C. Membranes were solubilized with 2% (w/v) DM for 1.5 h. Insoluble material was removed by ultracentrifugation and the receptors were immobilized by batch binding to TALON metal-affinity resin for 2 h. The resin was packed into a column and washed with steps of 8 and 30 mM imidazole in 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.15% (w/v) DM, and 0.3 μ M CP-376395 for a total of 15–20 column volumes before bound material was eluted with 200 mM imidazole. Receptors were further purified by gel filtration in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.15% (w/v) DM, and 0.3 μ M CP-376395. Preparations of CRF₁R 105 yielded more aggregated material than those of CRF₁R 76. To improve sample homogeneity, CRF₁R 105 was desalted into 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.15% (w/v) DM, 0.3 μ M CP-376395 and 5 mM EDTA after the metal-affinity step. Receptor purity was analysed using SDS-PAGE and receptor monodispersity was assayed by FSEC monitoring tryptophan fluorescence (Supplementary Fig. 10). Protein concentration was determined using the receptor's calculated extinction coefficient at 280 nm ($\epsilon_{280,calc} = 1.6 \text{ ml mg}^{-1} \text{ cm}^{-1}$).

Crystallization. CRF₁R was crystallized in lipidic cubic phase at 22.5 °C. The protein was concentrated to 20–30 mg ml^{−1} and mixed with monoolein (Nu-Check) supplemented with 10% (w/w) cholesterol (Sigma Aldrich) and 5 μ M CP-376395 using the twin-syringe method⁴⁵. The final protein:lipid ratio was 1:1.5 (w/w). 40–60 nl boluses were dispensed on 96-well glass bases and overlaid with 0.75 μ l precipitant solution. 20–30 μ m crystals of construct CRF₁R 76 were grown in 100 mM Na-citrate pH 5.5, 200 mM Li₂SO₄, 30% (v/v) polyethylene glycol 400, and 0.6 μ M CP-376395. A complete data set to 3.2 Å was obtained by combining diffraction data from multiple crystals. The crystals belonged to space group P6 and the data featured a 30% off-origin peak in a native Patterson map, indicating translational non-crystallographic symmetry (tNCS). Extensive trials to solve the structure by molecular replacement failed, most likely due to the presence of tNCS. We proposed that conformational flexibility in the connection between the receptor and T4L was the cause for the observed pseudo-symmetry and that shortening this part of the CRF₁R–T4L fusion would reduce flexibility and, hence, enable growth of a different crystal form. The resulting construct CRF₁R 105 (Fig. 1) crystallized in similar conditions to CRF₁R 76 and 20 μ m brick-shaped crystals of space group P22₁2₁ grew within 7–10 days (Supplementary Fig. 11). Crystals were flash-frozen in liquid nitrogen without additional cryoprotectant.

Diffraction data collection and processing. X-ray diffraction data were measured on a Pilatus 6M detector at Diamond Light Source beamline I24 using a 5 μ m \times 5 μ m microbeam. Crystals displayed isotropic diffraction to beyond 3.0 Å following exposure to an unattenuated beam for 7.5 s per degree of oscillation. Consequently, radiation damage was severe and wedges of typically only 2–3 degrees per crystal could be used for data merging. Data from individual crystals were integrated using XDS⁴⁶ and a complete data set was compiled using the data collection strategy option of the program Mosflm⁴⁷. Data merging and scaling was carried out with AIMLESS⁴⁸. Using the microdiffraction assembly method described previously⁴⁹ we were able to extend the resolution to 3.0 Å. Briefly, data from each crystal were split into bins of reflection observations corresponding to 1° of oscillation and scaled individually to a medium-resolution (4.3 Å) reference data set using XSCALE⁴⁶. Initially, as rejection criterion for reflections, the peak profile correlation threshold was set to zero and increased in increments of 1% until all reflection observations could be scaled with an R_{merge} lower than 14%. The resulting multi-record reflection file was then scaled and merged using AIMLESS. Data collection statistics for both methods are presented in Supplementary Table 6. For subsequent structure solution and refinement the data processed using the micro-diffraction assembly method were used.

Structure solution and refinement. The structure was solved by molecular replacement with Phaser^{48,50}, using two independent search models, T4L from the adenosine A_{2A} receptor structure (PDB ID 3EML) and a truncated version (TM helices only, no loops) of the dopamine D₃ receptor (PDB ID 3PBL). Solutions were found for two out of three T4L copies, which were subsequently fixed to locate three copies of the truncated receptor. Manual model building was done in COOT⁵¹ using sigma-A weighted $2F_o - F_c$, $F_o - F_c$ as well as a simulated-annealing composite omit maps calculated using PHENIX⁵². Initial refinement was carried out with REFMAC5^{48,53} using the maximum-likelihood restrained refinement protocol in combination with the jelly-body method and imposing tight non-crystallographic symmetry restraints. Later stages of the refinement were performed with PHENIX using a combination of simulated annealing, positional and individual isotropic B-factor refinement. The resulting model was then submitted to backbone torsion optimization followed by automated all-atom real-space refinement against a $2F_o - F_c$ electron density map, a method developed in ref. 54, resulting in improved stereochemistry and electron density

maps. Structure quality was assessed with MolProbity⁵⁵. With increasing quality of the model, weak electron density became visible for the first and last few residues of the missing copy of T4L in molecule C, revealing that the orientation of the T4L insertion relative to its corresponding receptor was significantly different from those observed in the other two molecules. However, only partial density was observed for the remaining parts of T4L (Supplementary Fig. 4) and T4L was therefore left unmodelled. Refinement statistics are presented in Supplementary Table 6.

Structure analysis. The root mean squared deviation (r.m.s.d.) analysis between CRF₁R and class A GPCRs (Supplementary Table 2) was performed using Maestro v. 9.3 (Schrödinger). D₃R (molecule A in PDB ID 3PBL) was superimposed onto CRF₁R molecule C using the C α -atoms of the following amino acid ranges comprising the cytoplasmic halves of TM1, TM2, TM4 and TM5 as well as entire TM3 (CRF₁/D₃R): 130–143/43–56 (TM1), 150–162/63–75 (TM2), 193–216/108–131 (TM3), 228–234/150–156 (TM4), 282–295/203–216 (TM5). TM6 and TM7, showing obvious structural differences, were excluded. Interhelical angles were calculated using interhlx (K. Yap, University of Toronto). Figures were prepared using PyMOL (Schrödinger).

41. Robertson, N. *et al.* The properties of thermostabilised G protein-coupled receptors (StaRs) and their use in drug discovery. *Neuropharmacology* **60**, 36–44 (2011).
42. Söding, J. Protein homology detection by HMM–HMM comparison. *Bioinformatics* **21**, 951–960 (2005).
43. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
44. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
45. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic cubic mesophases. *Nature Protocols* **4**, 706–731 (2009).
46. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
47. Leslie, A. G. W. & Powell, H. R. Processing diffraction data with Mosflm. *Evol. Methods Macromol. Crystallogr.* **245**, 41–51 (2007).
48. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
49. Hanson, M. A. *et al.* Crystal structure of a lipid G protein-coupled receptor. *Science* **335**, 851–855 (2012).
50. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
51. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
52. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
53. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
54. Haddadian, E. J. *et al.* Automated real-space refinement of protein structures using a realistic backbone move set. *Biophys. J.* **101**, 899–909 (2011).
55. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).

Receptor binding by an H7N9 influenza virus from humans

Xiaoli Xiong¹, Stephen R. Martin¹, Lesley F. Haire¹, Stephen A. Wharton², Rodney S. Daniels², Michael S. Bennett², John W. McCauley², Patrick J. Collins¹, Philip A. Walker¹, John J. Skehel¹ & Steven J. Gamblin¹

Of the 132 people known to have been infected with H7N9 influenza viruses in China, 37 died, and many were severely ill¹. Infection seems to have involved contact with infected poultry^{2,3}. We have examined the receptor-binding properties of this H7N9 virus and compared them with those of an avian H7N3 virus. We find that the human H7 virus has significantly higher affinity for α -2,6-linked sialic acid analogues ('human receptor') than avian H7 while retaining the strong binding to α -2,3-linked sialic acid analogues ('avian receptor') characteristic of avian viruses. The human H7 virus does not, therefore, have the preference for human versus avian receptors characteristic of pandemic viruses. X-ray crystallography of the receptor-binding protein, haemagglutinin (HA), in complex with receptor analogues indicates that both human and avian receptors adopt different conformations when bound to human H7 HA than they do when bound to avian H7 HA. Human receptor bound to human H7 HA exits the binding site in a different direction to that seen in complexes formed by HAs from pandemic viruses^{4,5} and from an aerosol-transmissible H5 mutant⁶. The human-receptor-binding properties of human H7 probably arise from the introduction of two bulky hydrophobic residues by the substitutions Gln226Leu and Gly186Val. The former is shared with the 1957 H2 and 1968 H3 pandemic viruses and with the aerosol-transmissible H5 mutant. We conclude that the human H7 virus has acquired some of the receptor-binding characteristics that are typical of pandemic viruses, but its retained preference for avian receptor may restrict its further evolution towards a virus that could transmit efficiently between humans, perhaps by binding to avian-receptor-rich mucins in the human respiratory tract⁷ rather than to cellular receptors.

Genetic analyses of the avian H7N9 influenza viruses, responsible for the zoonotic infections of humans in China since February 2013, indicate that the genes for the haemagglutinin (HA) and neuraminidase (NA) virus membrane glycoproteins are similar to those from viruses recently isolated from ducks in China (HA) and from wild birds and ducks from China and Korea (NA) (refs 3, 8 and A. Rambaut, http://epidemic.bio.ed.ac.uk/influenza_H7N9). The other six RNA molecules of their genomes derived from H9N2 viruses that have circulated widely in birds in the Far East and the Middle East since the late 1990s, causing sporadic human infections^{9–11}. We have used human H7N9 and avian H7N3 viruses grown in hens' eggs to determine their receptor binding affinity and specificity. On the basis of the reported sequences of HA we have also prepared baculoviruses containing complementary DNA for both H7 HAs and expressed them in insect cells to obtain HA proteins for crystallization. Our overall objective in these studies is to determine the structure of the HA from the H7N9 virus and compare its receptor-binding properties in particular, with those of viruses from the H1N1 pandemics of 1918 and 2009, the H2N2 pandemic of 1957, and the H3N2 pandemic of 1968, and with those that we recently determined for an aerosol-transmissible H5N1 mutant virus⁶.

We used biolayer interferometry to measure the binding of human H7N9 and avian H7N3 viruses to biosensors coated with human and avian receptor analogues (the terms 'human' and 'avian' receptor are used to refer to the sialosaccharides found to be abundant on airway epithelial cells in humans and to those in the avian gut, respectively) (Fig. 1). For comparison, in Supplementary Fig. 1a, we have included the binding curves for the H3 pandemic virus and the aerosol-transmissible H5 mutant virus, and in Supplementary Fig. 1b we show data for avian H3 and wild-type avian H5 viruses⁶. The data show that the human H7N9 virus has a substantially increased avidity for human receptor and a slightly decreased avidity for avian receptor by comparison with the avian H7 virus. The human virus thus has a modest preference for avian receptor binding (Fig. 1). In marked contrast, the aerosol-transmissible H5 mutant has much lower avidity than human H7 for human receptor and almost no binding affinity for avian receptor. Thus, although the transmissible H5 mutant shows the pandemic characteristic feature of preferring human to avian receptor, it achieves this by almost complete loss of avidity for avian receptor. The human H7 does not have the pandemic characteristic of preference for human receptor but it has

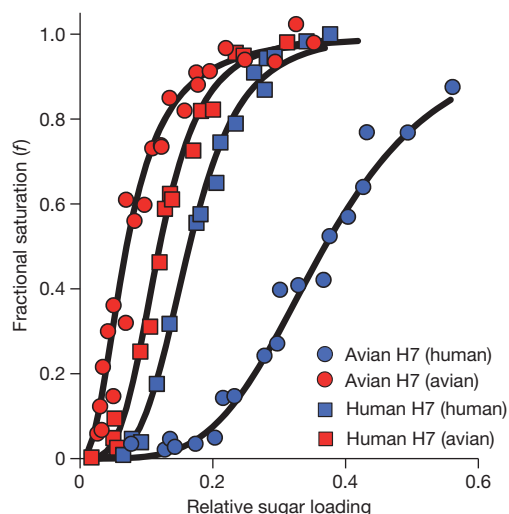


Figure 1 | Virus binding to human and avian receptors. Biolayer interferometry data for the binding of human H7N9 and avian H7N3 viruses to avian (α 2,3-SLN) and human (α 2,6-SLN) receptor analogues. 30-kDa polymers containing 20% mol sugar and 5% mol biotin linked to a polyacrylamide backbone were immobilized to different levels on streptavidin-coated biosensors. Data are plotted as fractional saturation of the sensor surface as a function of relative sugar loading (RSL) for a fixed virus concentration of 100 pM. The preference ratio of a virus is based on the relative avidities for human receptor versus avian receptor and its derivation is given in Methods. For the human H7N9 virus the preference ratio is 0.32 (that is, threefold tighter binding to avian receptor), whereas for pandemic H3 this ratio is 10 and for aerosol-transmissible H5 mutant it is 200 (Supplementary Fig. 1a).

¹MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW71AA, UK. ²WHO Collaborating Centre for Reference and Research on Influenza, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW71AA, UK.

much stronger avidity towards human (and avian) receptors than the aerosol-transmissible H5 mutant. The binding properties of the avian H7 virus (Fig. 1) are actually similar to those of the avian H3 virus (Supplementary Fig. 1b) and both are quite different from wild-type avian H5 which has very tight binding to avian receptors and virtually none to human receptors.

We have determined the crystal structures of human H7 HA and its complexes with human and avian receptor analogues (Fig. 2) and, for comparison, those of an avian H7 HA and its receptor complexes (Fig. 3). Crystallographic statistics are given in Supplementary Tables 1, 2 and 3. The receptor binding site is generally conserved in all HAs, but in the case of H7 it is notable that an insertion of two residues following residue 156 makes the 150-loop protrude into one edge of the binding site where it has the potential to influence receptor specificity¹². From the crystal structures it is evident that both human and avian receptors adopt different conformations when bound by the human H7 HA than when bound by avian H7 HA.

The human receptor complex with human H7 HA shows good electron density (Supplementary Fig. 2) for the sialic acid (Sia-1) and galactose-2 (Gal-2) moieties and partial density for the third sugar, N-acetyl glucosamine (NAG-3). The receptor takes an off-vertical trajectory out of the binding site (Fig. 2a). In contrast, the human receptor complex with avian H7 HA shows interpretable electron density for Sia-1 and Gal-2, but not for NAG-3, and the conformation is such that Gal-2 appears face-on (Fig. 2b) as opposed to edge-on in the human HA complex (Fig. 2a). The conformation of the receptor analogue

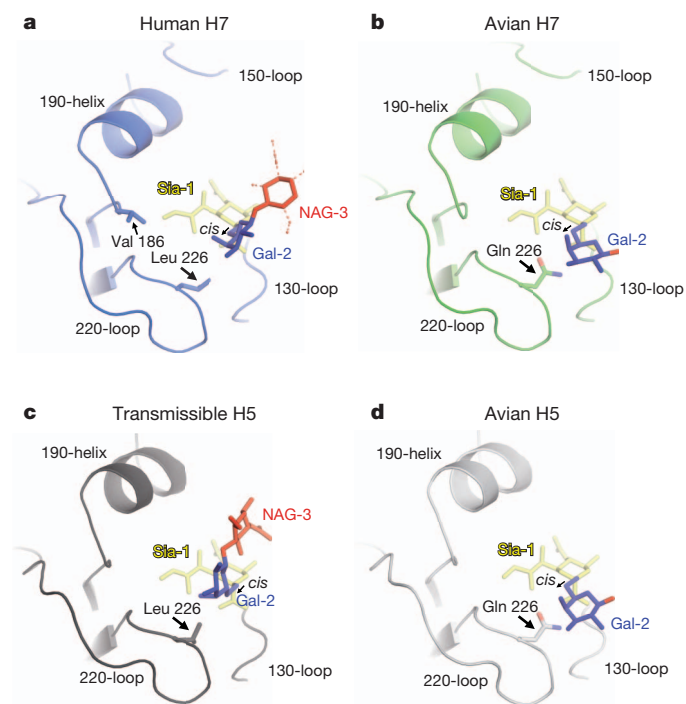


Figure 2 | Crystal structures of human receptor bound to human and avian HA. Ribbon representations of the receptor binding site, located at the membrane distal end of an HA monomer, showing the three conserved secondary structure elements of the site, 130-loop and 220-loop, and 190-helix. For H7 the 150-loop is longer and contributes to one edge of the site. **a**, Human receptor bound to human H7 HA with the side chains of the two hydrophobic substitutions (by comparison with avian H7) Gln226Leu and Gly186Val, shown in stick representation. Dotted lines are used to indicate the substituents of NAG-3 which are not defined by the electron density (Supplementary Fig. 1). **b**, Human receptor bound to avian H7 HA. In this case residue 226 is a glutamine, typical of avian HAs. **c**, Aerosol-transmissible H5 mutant HA⁶ bound to human receptor and the position of Leu 226. **d**, Avian H5 HA binds human receptor in a very similar manner to avian H7 (see **b**). For comparison, a matrix of sequence identities and coordinate differences for HA1 from H7, and the other HAs discussed, is presented in Supplementary Fig. 6.

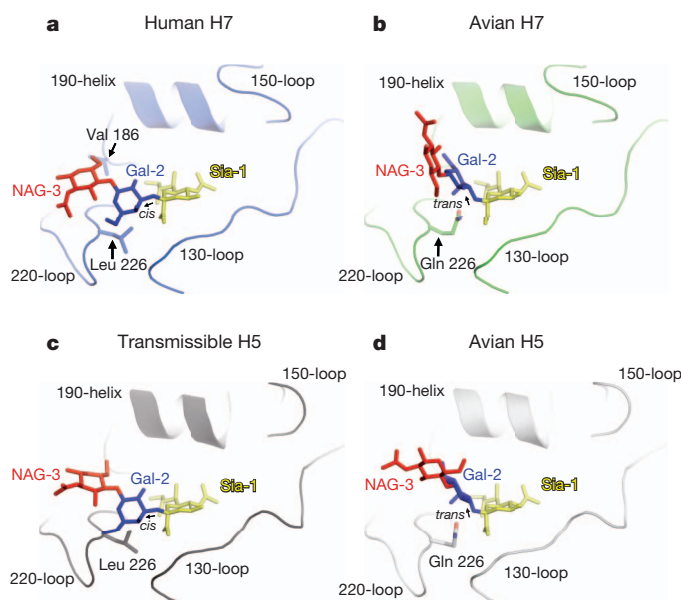


Figure 3 | Crystal structures of avian receptor bound to human and avian HA. Ribbon representations of the receptor binding site viewed after rotation of the molecule by 60° about a vertical axis relative to Fig. 2. **a**, Avian receptor bound to human H7 HA with a *cis* conformation about the Sia-1–Gal-2 glycosidic bond. **b**, Avian receptor bound to avian H7 HA, in this case with a *trans* conformation about the glycosidic bond. **c**, Aerosol-transmissible H5 mutant HA in complex with avian receptor and **d**, avian H5 HA in complex with avian receptor.

about the Sia-1–Gal-2 glycosidic bond is *cis* in both human and avian HA complexes, as it is in complexes formed by HAs of pandemic viruses and the aerosol-transmissible H5 mutant HA and wild-type avian H5 HA^{4–6}. The conformation of the human receptor bound to avian H7 HA is in fact very similar to that of human receptor bound to wild-type avian H5 HA (Fig. 2d). Analysis of the human receptor shows that there is a 45° difference in rotation around the Gal-2 C6–C5 bond between the human and avian H7 HA complexes. For comparison, between aerosol-transmissible H5 mutant and wild-type avian H5 HA complexes, the rotation is about 90°. As a consequence, Gal-2 and NAG-3 of the human receptor are 3–7 Å further away from the 190-helix in the human H7 HA complex than in the aerosol-transmissible H5 mutant complex (Fig. 2a, c).

There are two amino acid substitutions in the receptor-binding site between avian and human H7 HAs; the glutamine residue at position 226, characteristic of avian viruses, is replaced by a hydrophobic leucine and the glycine residue at position 186 in the avian HA changes to a bulkier hydrophobic valine residue in human H7 (residues are numbered throughout according to the H3 convention). The effect of these two substitutions is to make the part of the binding site where Gal-2 sits substantially more hydrophobic. The role of Leu 226 in human HAs is well documented^{14,15}; it generates a hydrophobic patch under the C6 of Gal-2 and widens the separation between the 220-loop and the 130-loop by about 1 Å. In the case of human H7 the leucine makes similar contacts with the carbons of the Gal-2 ring, but there is only a slight increase in the 220-loop–130-loop separation compared to avian H7. The potential role of Val 186 in human H7 is indicated by comparison with aerosol-transmissible H5 mutant HA, where the equivalent residue is an asparagine. Asn 186 is involved in a hydrogen-bond network with two water molecules that also involves the 3'-OH of Gal-2. The presence of valine at this position precludes the occupancy of the two water molecules and probably accounts for the fact that Gal-2 sits further out of the site than is the case in the aerosol-transmissible H5 mutant HA complex.

The avian receptor complex of human H7 HA shows good electron density for Sia-1, Gal-2 and NAG-3 of the receptor, as does the avian receptor complex of avian H7 HA (Supplementary Fig. 2). The avian

receptor adopts a *cis* configuration about the Sia-1–Gal-2 glycosidic bond when bound to human H7 HA (Fig. 3a), but a *trans* arrangement when bound to avian H7 HA (Fig. 3b)—which is typical of avian receptors binding to avian HAs. The avian receptor bound to human H7 adopts a similar arrangement to that seen for the aerosol-transmissible H5 mutant complex, although Gal-2 and NAG-3 are rotated slightly differently (Fig. 3a, c). As discussed above, the introduction of the bulky hydrophobic leucine residue at position 226 alters the binding environment leading to the switch from a *trans* to *cis* configuration. The conformation of avian receptor bound to avian H7 is very similar to that in other avian receptor–avian HA complexes¹³ for Sia-1 and Gal-2, but there is a 65° rotation of NAG-3 relative to Gal-2 that would mean an extended sialosaccharide chain would exit the avian HA receptor binding site in a somewhat different direction to that normally seen for avian receptor complexes (Fig. 3b, d).

From our comparisons of the structures of human and avian H7 HAs and the complexes that each forms with human and avian receptor analogues, together with our observations on receptor binding affinity and specificity, we deduce a structural basis for the role of the H7 HA in human infection and consider its potential contribution to virus transmissibility and to the severity of infection. Like the HAs from H2 and H3 pandemic viruses, the human H7 HA contains the Gln226Leu substitution that provides a non-polar binding site for C-6 and the ring carbons of Gal-2 of the human receptor. The substitution Gly186Val, which has been reported to influence receptor-binding specificity in European H7 viruses¹⁴, also increases the hydrophobicity of the binding site to favour interactions with human receptor. As a result the human receptor adopts a conformation in which NAG-3 extends towards the top of the site, but in a less vertical position than in similar complexes formed by either H1, H2 or H3 pandemic virus HAs^{5,15}. This conformation is, however, markedly different from that adopted by human receptor in complex with avian H7 HA (Fig. 2).

A particular structural feature of H7, H10 and H15 (the H7 clade) receptor binding sites¹⁶, by comparison with those of all other HAs, results from an insertion of two residues after position 156, which causes the 150-loop to protrude into the site (Fig. 2). This protrusion may also influence the orientation of the human receptor bound to human H7 HA, because binding of human receptor in the more vertical conformations (observed in complexes with the HAs of H1, H2 and H3 pandemic viruses) would be sterically restricted (Supplementary Fig. 3 and ref. 12).

In close similarity to the HAs of pandemic viruses, and to the HA of the aerosol-transmitted H5 mutant virus^{6,17}, the human H7 HA binds avian receptor in a different conformation from that observed in the complex formed by avian H7 HA with avian receptor. However, although human H7 virus shows increased affinity for human receptor, by comparison with avian H7 virus, it retains a similar affinity for avian receptor. Because mucins in the human respiratory tract are rich in α 2,3-linked sialic acid (avian receptor)⁷, this retained affinity of human H7 virus for avian receptors and the interactions with mucins that could result would restrict its binding to human cellular receptors and limit infection.

In addition to its role in receptor binding, a number of other properties of HA could influence transmission of the H7N9 virus: (1) HA is the influenza virus membrane fusion protein that mediates entry of the genome-transcriptase complex into cells during infection; (2) because both HA (receptor-binding protein) and NA (receptor-destroying enzyme) glycoproteins of the influenza virus membrane recognize sialic acid, a balance between their activities is required for efficient virus replication^{18,19}.

In relation to HA function in membrane fusion, there are no obvious sequence differences between the human and avian H7 HAs in the regions of HA precursor cleavage or the fusion subdomain that could influence virus pathogenicity^{20–22}. However, the pH at which conformational changes in HA occur, that are required for membrane fusion, is higher for human H7 HA than for HAs of the H7N3 virus and of

pandemic viruses (pH 6.3 versus pH 5.6, Supplementary Fig. 4 and refs 21, 23). This indicates that the human H7 HA is comparatively less thermostable²⁴, and may require mutation to increase its stability, in a similar way to that achieved by the Thr318Ile substitution in the aerosol-transmissible H5 mutant HA^{6,25}, before it would be able to transmit efficiently between mammals.

Finally, a balance of HA and NA activities is required for efficient infection. In poultry H7 viruses, from which the human H7N9 virus derives, deletions that shorten the NA stalk and decrease its enzyme activity are commonly observed¹⁴. In our biolayer interferometry studies of virus receptor binding we were also able to estimate the NA activity of the human H7N9 virus, by comparison with pandemic H3N2 virus (Supplementary Fig. 5). For both avian and human receptor analogues the human H7N9 virus NA activity was considerably lower. Because the receptor-binding properties of the human H7N9 virus are likely to be different from the avian H7N9 virus, and the NA activity is clearly low for human receptor, a combination of increased NA activity and/or decreased HA affinity for human receptor is likely to be required for efficient transmission of the H7N9 virus in mammals.

METHODS SUMMARY

The HA genes from two viruses. A/Anhui/1/2013 (H7N9) (human H7) and A/turkey/Italy/214845/2002 (H7N3) (avian H7) were sub-cloned into a modified pAcGP67A vector which carries a TEV protease site, a trimerization foldon and a His-tag. The initial human H7 HA clone we used differed at two residues from the previously published sequence (Leu 20 and Thr 135), and a second clone was also produced for structural studies that contained the published sequence (Val 20 and Ala 135). Proteins were expressed in Sf-9 cells and purified as described before²⁶. Both human and avian HAs were crystallized from 0.1 M PIPES pH 7.0, 2.3 M ammonium sulphate, 1% PEG 400. Crystals were cryoprotected with crystallization solution plus 20% ethylene glycol and HA–receptor complexes were prepared by soaking HA crystals in cryoprotected crystallization solution supplemented with 100 mM receptor analogues α 2,3- and α 2,6-linked sialyl lactosamine sugars (α 2,3-SLN and α 2,6-SLN, respectively). Diffraction data were collected at 100 K at the Diamond synchrotron, processed in the XIA2 pipeline. Structures were built with Coot²⁷ and refined with Refmac²⁸. The receptor complex structures obtained from both clones are very similar and are presented in Supplementary Fig. 2.

Biolayer interferometry. Influenza viruses were grown in hens' eggs and purified according to standard protocols²⁹. Virus binding to defined receptor analogues was measured on an Octet RED biolayer interferometer (Pall ForteBio). Biotinylated α 2,3-SLN and α 2,6-SLN were purchased from Lectinity Holding. Binding of viruses (at 100 pM) was measured at 25 °C in a 30–50 min association step. All solutions also contained 10 μ M oseltamivir carboxylate (Roche) and 10 μ M zanamivir (GSK) to prevent cleavage of the receptor analogues by the viral neuraminidase. The (relative) amount of virus bound to the biosensor at different relative sugar loadings was calculated from the amplitude of the response at the end of the association step.

Full Methods and any associated references are available in the online version of the paper.

Received 27 May; accepted 13 June 2013.

Published online 20 June 2013.

1. World Health Organization. Number of confirmed human cases of avian influenza A(H7N9) reported to WHO Report 8 - data in WHO/HQ as of 30 May 2013, 15:45 GMT+1. http://www.who.int/influenza/human_animal_interface/influenza_h7n9/08_ReportWebH7N9Number.pdf (2013).
2. Chen, Y. *et al.* Human infections with the emerging avian influenza A H7N9 virus from wet market poultry: clinical analysis and characterisation of viral genome. *Lancet* **381**, 1916–1925 (2013).
3. Gao, R. *et al.* Human infection with a novel avian-origin influenza A (H7N9) virus. *N. Engl. J. Med.* **368**, 1888–1897 (2013).
4. Eisen, M. B., Sabesan, S., Skehel, J. J. & Wiley, D. C. Binding of the influenza A virus to cell-surface receptors: structures of five hemagglutinin–sialyloligosaccharide complexes determined by X-ray crystallography. *Virology* **232**, 19–31 (1997).
5. Liu, J. *et al.* Structures of receptor complexes formed by hemagglutinins from the Asian influenza pandemic of 1957. *Proc. Natl Acad. Sci. USA* **106**, 17175–17180 (2009).
6. Xiong, X. *et al.* Receptor binding by a ferret-transmissible H5 avian influenza virus. *Nature* **497**, 392–396 (2013).
7. Couceiro, J. N., Paulson, J. C. & Baum, L. G. Influenza virus strains selectively recognize sialyloligosaccharides on human respiratory epithelium; the role of the host cell in selection of hemagglutinin receptor specificity. *Virus Res.* **29**, 155–165 (1993).

8. Kageyama, T. *et al.* Genetic analysis of novel avian A(H7N9) influenza viruses isolated from patients in China, February to April 2013. *Eurosurveillance* **18**, 20453 (2013).
9. Butt, K. M. *et al.* Human infection with an avian H9N2 influenza A virus in Hong Kong in 2003. *J. Clin. Microbiol.* **43**, 5760–5767 (2005).
10. Lin, Y. P. *et al.* Avian-to-human transmission of H9N2 subtype influenza A viruses: relationship between H9N2 and H5N1 human isolates. *Proc. Natl Acad. Sci. USA* **97**, 9654–9658 (2000).
11. Peiris, M. *et al.* Human infection with influenza H9N2. *Lancet* **354**, 916–917 (1999).
12. Russell, R. J., Stevens, D. J., Haire, L. F., Gamblin, S. J. & Skehel, J. J. Avian and human receptor binding by hemagglutinins of influenza A viruses. *Glycoconj. J.* **23**, 85–92 (2006).
13. Ha, Y., Stevens, D. J., Skehel, J. J. & Wiley, D. C. X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs. *Proc. Natl Acad. Sci. USA* **98**, 11181–11186 (2001).
14. Gambaryan, A. S. *et al.* Receptor-binding profiles of H7 subtype influenza viruses in different host species. *J. Virol.* **86**, 4370–4379 (2012).
15. Gamblin, S. J. *et al.* The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* **303**, 1838–1842 (2004).
16. Russell, R. J. *et al.* H1 and H7 influenza haemagglutinin structures extend a structural classification of haemagglutinin subtypes. *Virology* **325**, 287–296 (2004).
17. Zhang, W. *et al.* An airborne transmissible avian influenza H5 hemagglutinin seen at the atomic level. *Science* <http://dx.doi.org/10.1126/science.1236787> (2 May 2013).
18. Baigent, S. J. & McCauley, J. W. Influenza type A in humans, mammals and birds: determinants of virus virulence, host-range and interspecies transmission. *BioEssays* **25**, 657–671 (2003).
19. Wagner, R., Matrosovich, M. & Klenk, H. D. Functional balance between haemagglutinin and neuraminidase in influenza virus infections. *Rev. Med. Virol.* **12**, 159–166 (2002).
20. Garten, W. & Klenk, H. D. Understanding influenza virus pathogenicity. *Trends Microbiol.* **7**, 99–100 (1999).
21. Skehel, J. J. & Wiley, D. C. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu. Rev. Biochem.* **69**, 531–569 (2000).
22. Steinhauer, D. A. Role of hemagglutinin cleavage for the pathogenicity of influenza virus. *Virology* **258**, 1–20 (1999).
23. Daniels, R. S. *et al.* Fusion mutants of the influenza virus hemagglutinin glycoprotein. *Cell* **40**, 431–439 (1985).
24. Ruigrok, R. W. *et al.* Conformational changes in the hemagglutinin of influenza virus which accompany heat-induced fusion of virus with liposomes. *Virology* **155**, 484–497 (1986).
25. Imai, M. *et al.* Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature* **486**, 420–428 (2012).
26. Lin, Y. P. *et al.* Evolution of the receptor binding properties of the influenza A(H3N2) haemagglutinin. *Proc. Natl Acad. Sci. USA* **109**, 21474–21479 (2012).
27. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
28. Collaborative Computational Project, number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
29. Skehel, J. J. & Schild, G. C. The polypeptide composition of influenza A viruses. *Virology* **44**, 396–408 (1971).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge Y. Shu for sharing information and samples on recent H7N9 virus infections. We are grateful to staff at the Diamond Light Source Synchrotron for assistance and beamline access under proposal 7707 and the staff of the NIMR Large Scale laboratory. This work was funded by the Medical Research Council through programmes U117584222, U117570592, U117585868 and U117512723.

Author Contributions X.X., S.R.M., L.F.H., S.A.W., R.S.D., M.S.B., J.W.M., P.J.C., P.A.W., J.J.S. and S.J.G. all performed experiments and contributed to the writing of the manuscript.

Author Information Structural data have been deposited with the Protein DataBank with accession codes: 4BSA, 4BSB, 4BSC, 4BSD, 4BSE, 4BSF, 4BSG, 4BSH and 4BSI. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.J.G. (sgambli@nimr.mrc.ac.uk) or J.J.S. (skeheljj@nimr.mrc.ac.uk).

METHODS

The HA genes from two viruses. A/Anhui/1/2013 (H7N9) (human H7) and A/turkey/Italy/214845/2002 (H7N3) (avian H7) were sub-cloned into a modified pAcGP67A vector which carries a TEV protease site, a trimerization foldon and a His-tag. The initial human H7 HA clone we used differed at two residues from the previously published sequence (Leu 20 and Thr 135), and a second clone was also produced for structural studies that contained the published sequence (Val 20 and Ala 135). Proteins were expressed in Sf-9 cells and purified as described before²⁶. Both human and avian HAs were crystallized from 0.1 M PIPES pH 7.0, 2.3 M ammonium sulphate, 1% PEG 400. Crystals were cryoprotected with crystallization solution plus 20% ethylene glycol and HA–receptor complexes were prepared by soaking HA crystals in cryoprotected crystallization solution supplemented with 100 mM receptor analogues α 2,3- and α 2,6-linked sialyl lactosamine sugars (α 2,3-SLN and α 2,6-SLN, respectively). Diffraction data were collected at 100 K at the Diamond synchrotron, processed in the XIA2 pipeline. Structures were built with Coot²⁷ and refined with Refmac²⁸. The receptor complex structures obtained from both clones are very similar and are presented in Supplementary Fig. 2.

Biolayer interferometry. Influenza viruses were grown in hens' eggs and purified according to standard protocols²⁹.

Virus binding to defined receptor analogues was measured on an Octet RED biolayer interferometer (Pall ForteBio). Biotinylated α 2,3- and α 2,6-linked sialyl lactosamine sugars (α 2,3-SLN and α 2,6-SLN, respectively) were purchased from Lectinity Holding. These were approximately 30 kDa polymers containing 20% mol sugar and 5% mol biotin linked to a polyacrylamide backbone. The polymers were immobilized on streptavidin biosensors (Pall ForteBio) at concentrations ranging from 0.01 to 1.5 $\mu\text{g ml}^{-1}$. The relative sugar loading (RSL) of the biosensor was calculated from the amplitude of the response at the end of the 5–10 min loading step. The maximum response at complete saturation was ~ 0.6 nm.

Binding of viruses (at 100 pM) was measured at 25 °C in a 30–50 min association step. The buffer was 10 mM HEPES (pH 7.4), 150 mM NaCl, 3 mM EDTA and 0.005% Tween-20. All solutions also contained 10 μM oseltamivir carboxylate (Roche) and 10 μM zanamivir (GSK) to prevent cleavage of the receptor analogues by the viral neuraminidase. The drugs were omitted in some experiments designed to investigate the ability of the neuraminidase to cleave the sialic acids from the sensor surface.

The (relative) amount of virus bound to the biosensor at different relative sugar loadings was calculated from the amplitude of the response at the end of association step. These measured amplitudes were normalized by dividing by the maximum response (typically 5–6 nm) and this normalized response was plotted as a function of the relative sugar loading (see Fig. 1 main text). These normalized virus binding response curves report the fractional saturation of the sensor surface (f) and smooth lines through the curves were generated by fitting the data to a simple variant of the Hill equation:

$$f = \frac{[\text{RSL}]^n}{[\text{RSL}_{0.5}]^n + [\text{RSL}]^n}$$

where RSL is the relative sugar loading, $\text{RSL}_{0.5}$ is the relative sugar loading at half saturation ($f = 0.5$), and n is a Hill coefficient. The fractional saturation of the sensor surface (f) can be related to an apparent equilibrium dissociation constant for virus binding, $K_{\text{d(Virus)}}$ in the normal way:

$$f = \frac{[\text{Virus}]}{[\text{Virus}] + K_{\text{d(Virus)}}}$$

So that $K_{\text{d(Virus)}} = \frac{[\text{Virus}](1-f)}{f}$ and $K_{\text{d(Virus)}}$ can then be calculated as a function

of the relative sugar loading for f values in the range 0.05 to 0.95. These $K_{\text{d(Virus)}}$ values were used to calculate the preference ratio for human/avian receptors. For example, for the human H7 both receptor analogues have f values in the range 0.05 to 0.95 for RSL values in the range 0.1 to 0.25. Over this range $K_{\text{d(Virus)}}$ for the avian receptor changes from ~ 200 pM (at $\text{RSL} = 0.1$) to ~ 6 pM (at $\text{RSL} = 0.25$) whereas $K_{\text{d(Virus)}}$ for the human receptor changes from ~ 750 pM (at $\text{RSL} = 0.1$) to ~ 18 pM (at $\text{RSL} = 0.25$). We therefore estimate that the preference ratio calculated over this range of RSL values is 0.32 ± 0.04 . In the same way, the human receptor:avian receptor preference ratio for the H3 pandemic virus = 10:1, the preference ratio for the aerosol-transmissible H5 mutant virus = 200:1, and the preference ratio for the avian H5 virus = 5×10^{-7} .

All measurements with human H7 virus were made under appropriate biological safety conditions.

Biological features of novel avian influenza A (H7N9) virus

Jianfang Zhou^{1*}, Dayan Wang^{1*}, Rongbao Gao^{1*}, Baihui Zhao^{2*}, Jingdong Song¹, Xian Qi³, Yanjun Zhang⁴, Yonglin Shi⁵, Lei Yang¹, Wenfei Zhu¹, Tian Bai¹, Kun Qin¹, Yu Lan¹, Shumei Zou¹, Junfeng Guo¹, Jie Dong¹, Libo Dong¹, Ye Zhang¹, Hejiang Wei¹, Xiaodan Li¹, Jian Lu¹, Liqi Liu¹, Xiang Zhao¹, Xiyun Li¹, Weijuan Huang¹, Leying Wen¹, Hong Bo¹, Li Xin¹, Yongkun Chen¹, Cuilin Xu¹, Yuquan Pei⁶, Yue Yang⁶, Xiaodong Zhang⁶, Shiwen Wang⁴, Zijian Feng⁷, Jun Han⁷, Weizhong Yang⁷, George F. Gao⁷, Guizhen Wu¹, Dexin Li¹, Yu Wang⁷ & Yuelong Shu¹

Human infection associated with a novel reassortant avian influenza H7N9 virus has recently been identified in China¹. A total of 132 confirmed cases and 39 deaths have been reported². Most patients presented with severe pneumonia and acute respiratory distress syndrome^{3,4}. Although the first epidemic has subsided, the presence of a natural reservoir and the disease severity highlight the need to evaluate its risk on human public health and to understand the possible pathogenesis mechanism. Here we show that the emerging H7N9 avian influenza virus poses a potentially high risk to humans. We discover that the H7N9 virus can bind to both avian-type (α 2,3-linked sialic acid) and human-type (α 2,6-linked sialic acid) receptors. It can invade epithelial cells in the human lower respiratory tract and type II pneumonocytes in alveoli, and replicated efficiently in *ex vivo* lung and trachea explant culture and several mammalian cell lines. In acute serum samples of H7N9-infected patients, increased levels of the chemokines and cytokines IP-10, MIG, MIP-1 β , MCP-1, IL-6, IL-8 and IFN- α were detected. We note that the human population is naive to the H7N9 virus, and current seasonal vaccination could not provide protection.

In February 2013, a novel avian influenza H7N9 virus emerged in China and was found to infect humans¹. As of 31 May 2013, a total of 132 human infections and 39 fatal cases have been identified². Most cases have occurred among elders who were considered to be at increased risk for influenza complications because of age immunity and/or underlying medical conditions⁴. In contrast to other previous human H7 infections that mainly manifested mild illness or conjunctivitis, most H7N9 patients were critically ill and presented severe pneumonia and acute respiratory distress syndrome^{1,3,4}. Sequence analysis of the avian influenza A (H7N9) virus indicates that it possesses several genetic features contributing to its ability to infect humans, and raises concerns about the pandemic potential¹. Therefore, to evaluate its risk on human public health and understand the possible pathogenesis mechanism, we investigated the biological features of the H7N9 virus and the host immunity of infected patients as well as the general human population.

It is generally accepted that haemagglutinin-receptor-binding preference to α 2,6-linked (human-type) sialylated glycans is the initial key step for a novel influenza-virus-causing pandemic. First, we tested the receptor-binding property of the novel H7N9 viruses (A/Shanghai/1/2013 (SH1), A/Shanghai/2/2013 (SH2), and A/Anhui/1/2013 (AH1)) using glycan dose-dependent binding assays as described previously⁵. We discovered that the viruses bound to both avian-type (α 2,3) and human-type (α 2,6) receptors, whereas avian influenza H5N1 (A/Anhui/01/2005) and pandemic H1N1 (A/California/07/2009) viruses preferentially bound to α 2,3 and α 2,6 receptors respectively, as expected (Fig. 1). The enhanced

α 2,6-binding of H7N9 may attribute to single or multiple substitutions in haemagglutinin including Ala138Ser in SH1 or Gly186Val and Gln226Leu in both SH2 and AH1, as reported elsewhere^{6–11}. The presence of alanine at residue 160 in haemagglutinin of all three viruses resulted in a glycosylation loss and had been observed to increase the affinity of the H5N1 virus for human-type receptors^{12,13}. Furthermore,

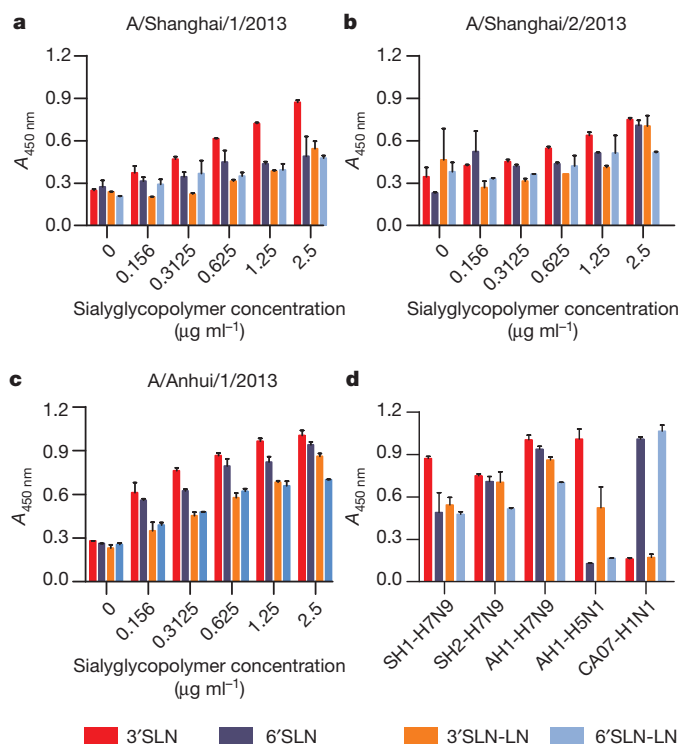


Figure 1 | Receptor binding properties. a–c, Dose-dependent binding of A/Shanghai/1/2013 (a), A/Shanghai/2/2013 (b) and A/Anhui/1/2013 (c) viruses to both α 2,3 and α 2,6 glycans at the concentration of 0, 0.156, 0.3125, 0.625, 1.25 and 2.5 μ g ml⁻¹ sialylglycopolymers. d, The binding pattern of H7N9, H5N1 (A/Anhui/01/2005) and pandemic H1N1 (A/California/07/2009) viruses were exhibited at 2.5 μ g ml⁻¹. H5N1 data were obtained previously, as H5N1 and H7N9 viruses cannot be simultaneously handled owing to biosafety reasons. Results denote mean \pm s.e.m. from two separate assays performed in duplicate. 3'SLN, Neu5Ac α 2-3Gal β 1-4GlcNAc; 3'SLN-LN, Neu5Ac α 2-3(Gal β 1-4GlcNAc β 1-3)₂; 6'SLN, Neu5Ac α 2-6Gal β 1-4GlcNAc; 6'SLN-LN, Neu5Ac α 2-6(Gal β 1-4GlcNAc β 1-3)₂.

¹National Institute for Viral Disease Control and Prevention, China CDC, Key Laboratory for Medical Virology, National Health and Family Planning Commission, Beijing 102206, China. ²Shanghai Municipal Disease Control and Prevention, Shanghai 200336, China. ³Jiangsu Provincial Disease Control and Prevention, Nanjing 210009, China. ⁴Zhejiang Provincial Disease Control and Prevention, Hangzhou 310051, China. ⁵Anhui Provincial Disease Control and Prevention, Hefei 230601, China. ⁶Key Laboratory of Carcinogenesis and Translational Research, Peking University Cancer Hospital, Beijing 100142, China. ⁷Chinese Center for Disease Control and Prevention, Beijing 102206, China.

*These authors contributed equally to this work.

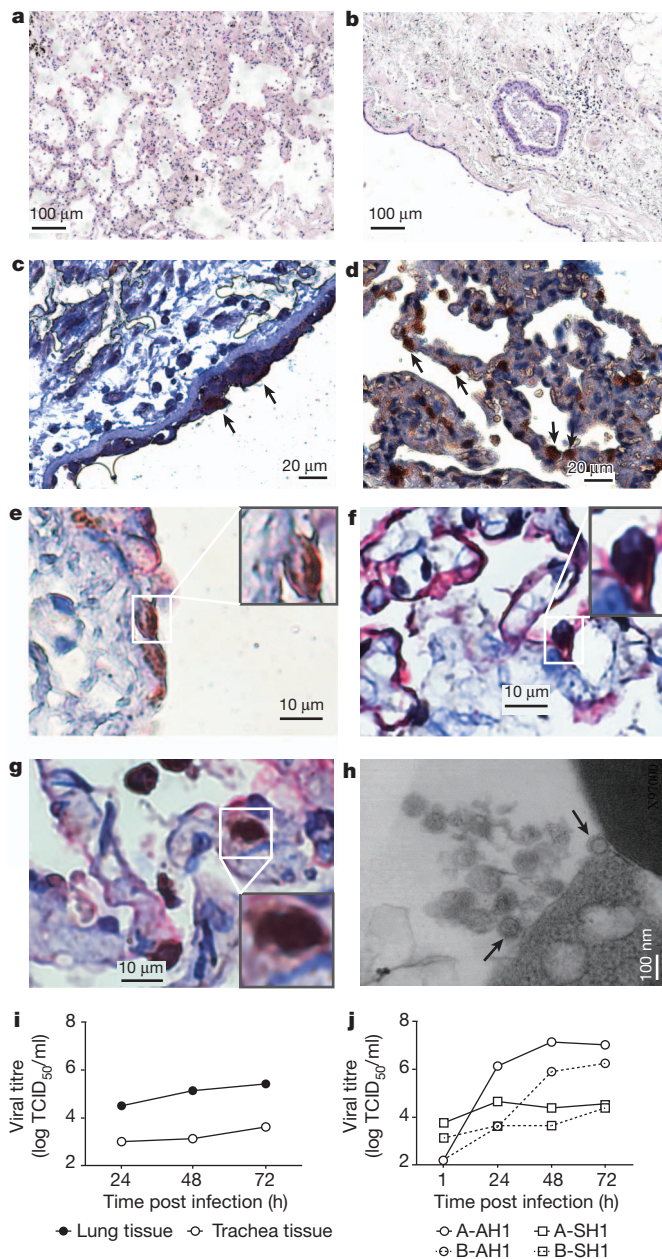


Figure 2 | Ex vivo infection of trachea or lung tissue with A/Anhui/1/2013 (AH1) or A/Shanghai/1/2013 (SH1). **a, b**, Trachea (**a**) and lung (**b**) explants without infection. **c, d**, Influenza nucleoprotein (brown, arrows) detected in trachea (**c**) and lung (**d**). **e–g**, Nucleoproteins in trachea epithelia/pneumocytes (**e, f**; red, cytokeratin staining) and type II pneumocytes (**g**) (red, surfactant staining). **h**, Extracellular viral particles (arrows) in AH1-infected lung detected by transmission electron microscopy. **i**, Viral replication in AH1-infected lung/trachea tissues from one donor (filled/empty circle, respectively). **j**, Viral yields of lung fragments from two additional donors (A/B, solid/dotted line, respectively) infecting AH1 (circle) and SH1 (square). Original magnifications: $\times 10$ (**a, b**), $\times 40$ (**c, d**) and $\times 100$ (**e–g**).

H7N9 viruses bound to long $\alpha 2,6$ glycans ($\alpha 2-6$ SLN-LN) efficiently (Fig. 1). Long $\alpha 2-6$ oligosaccharides with an umbrella-like topology, which are rich in human and swine respiratory epithelial cells, are involved in the adaptation and infection of influenza viruses in these hosts^{5,14,15}. The increased preference of the H7N9 virus for human-type receptors may result in its higher cross-species transmissibility than the H5N1 viruses, which prefer avian-type receptors. This receptor specificity had been considered to be responsible for the major species barrier preventing H5N1 from efficient infection and transmission in humans.

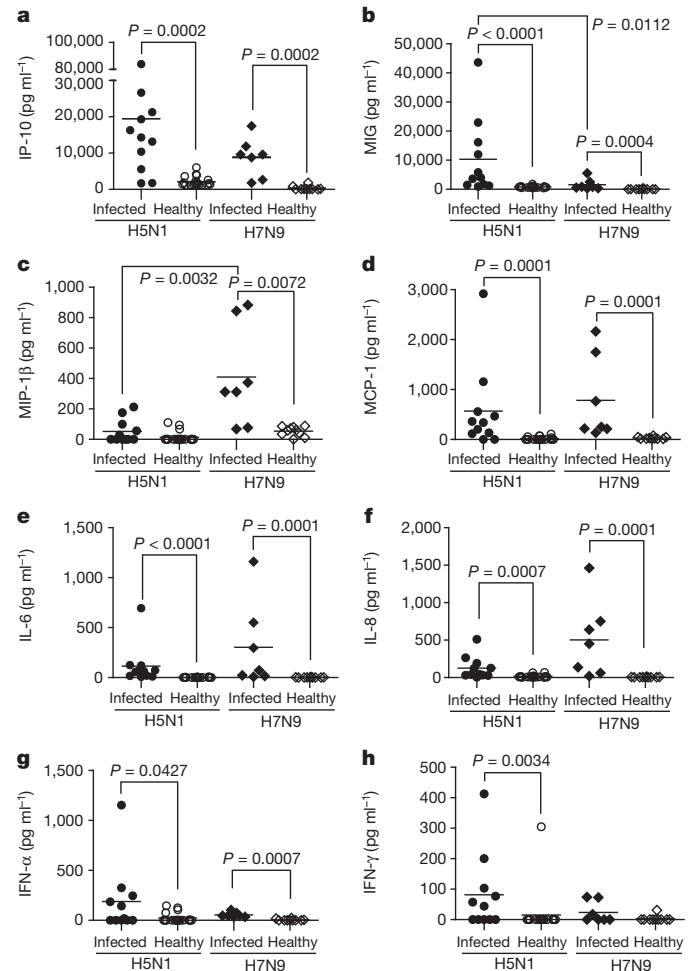


Figure 3 | Sera chemokine/cytokine levels of acutely infected patients by avian H7N9 or H5N1 viruses. **a–h**, Sera concentrations of the chemokines and cytokines IP-10 (**a**), MIG (**b**), MIP-1 β (**c**), MCP-1 (**d**), IL-6 (**e**), IL-8 (**f**), IFN- α (**g**) and IFN- γ (**h**) were measured using cytometric bead-array assays according to the manufacturers' instructions. Results are presented as mean \pm s.e.m. from the specimens of 11 H5N1, 7 H7N9 and 30 close contacts tested in duplicate. *P* values are shown on the figure; differences between groups were tested using non-parametric Kruskal–Wallis analysis of variance (ANOVA) with Dunn's post-hoc test for multiple comparisons.

Human upper respiratory tissues and trachea contain mainly $\alpha 2,6$ receptors, whereas lung tissues contain a mixture of $\alpha 2,3$ - and $\alpha 2,6$ -linked sialic acid^{5,14}. To assess the effect of dual-receptor binding of the H7N9 virus to human respiratory tracts, sections of tracheal and lung explants were infected. We found that the epithelial cells in the human lower respiratory tract and type II pneumocytes in the alveoli were susceptible for H7N9 virus infections, as evidenced by the expression of viral nucleoprotein (Fig. 2c–g). Extracellular viral particles were also observed in infected lung tissues at 24 hours post infection (h.p.i.) by transmission electron microscopy (Fig. 2h). Productive viral infection in the trachea and lung was demonstrated by increased viral yields in culture supernatants. Notably, the virus titre in lung tissues was about tenfold higher than that in tracheal tissues (Fig. 2i), which was consistent with the results of nucleoprotein staining (Fig. 2c, d). Efficient replication of the H7N9 virus in type II pneumocytes may cause lung function to deteriorate¹⁶, as seen with the H5N1 virus¹⁷. Lower viral yields in the trachea may be associated with the inefficient human-to-human transmission. Although AH1 and SH1 show common high binding preferences to $\alpha 2,3$ glycans, a more productive replication of AH1 in lung tissues was observed, with a peaking titre around 100-fold higher than that of SH1 (Fig. 2j). This indicated that viral replications in respiratory tissues are not solely determined by receptor-binding

Table 1 | Cross-reactive antibody response against the H7N9 virus

Age group	Subjects	Age distribution (years)	Antigen	Increase in antibody titre by a factor of ≥ 4 (%)	Geometric mean titre		Titre ≥ 40 (%)		Titre ≥ 160 (%)	
					Before vaccination	After vaccination	Before vaccination (%)	After vaccination (%)	Before vaccination (%)	After vaccination (%)
Children	30	3–5	CA07	76.7	20.5	272.2	43.3	80.0	10.0	73.3
			Vic361	36.7	55.3	152.8	73.3	93.3	16.7	60.0
			HB158	86.7	36.5	432.1	63.3	96.7	13.3	83.3
			AH1	0	<5	<5	0	0	0	0
Adults	30	18–59	CA07	70.0	14.8	156.3	26.7	76.7	0	60.0
			Vic361	53.3	9.8	38.2	13.3	70.0	0.0	13.3
			HB158	80.0	28.9	259.9	53.3	96.7	10.0	76.7
			AH1	0	<5	<5	0	0	0	0
Older people	30	≥ 60	CA07	46.7	10.5	52.8	10.0	53.3	0	30.0
			Vic361	40.0	25.8	68.1	40.0	90.0	0.0	30.0
			HB158	60.0	33.2	211.1	63.3	93.3	6.7	53.3
			AH1	0	<5	<5	0	0	0	0

Cross-reactive antibody response against the novel H7N9 virus in paediatric and adult recipients of 2012–2013 seasonal trivalent inactivated influenza vaccines. All children received two doses of vaccine, with a 1-month interval. The composition of the trivalent vaccine were A/Christchurch/16/2010 (NIB-74xp) (A/California/07/2009-like), A/Victoria/361/2011 (H3N2) IVR-165 and B/Hubei-Wujiagang/158/2009. AH1, A/Anhui/1/2013 (H7N9); CA07, A/California/07/2009 (H1N1); HB158, B/Hubei-Wujiagang/158/2009; Vic361, A/Victoria/361/2011 (H3N2).

properties, but that other factors, such as viral polymerase and host factors, may also have a role. Moreover, H7N9 isolates demonstrated advanced growth capabilities in human, porcine and canine cell lines that were comparable to that of the pandemic 2009 H1N1 virus (Supplementary Fig. 2).

The dysfunction of the chemokine and cytokine response is considered to be associated with severe human H5N1 infection^{18,19}. Therefore, we measured the concentrations of chemokines and cytokines in acute sera from seven patients infected with H7N9 by cytometric beads array (detailed sampling information shown in Supplementary Table 3). The levels of IP-10, MIG, MIP-1 β , MCP-1, IL-6, IL-8 and IFN- α were significantly higher in patients with H7N9 than in healthy subject controls (Fig. 3). No significant differences were detected between patients with H7N9 and H5N1, with the exception of the MIG and MIP-1 β levels. The MIG response was significantly lower in patients with H7N9 than with H5N1, whereas the MIP-1 β response was significantly higher in patients with H7N9 than with H5N1 (Fig. 3). The hypercytokinaemia in patients with H7N9 may be one of the important contributors to clinical severity, as seen in H5N1 infections²⁰. Notably, most of the severe H7N9 cases were in old adults (≥ 60 years), which might be due to the enhanced expression of chemokine receptor related to ageing or chronic diseases, as observed previously^{21–24}.

Pre-existing immunity in the general population is one of the determining factors for the susceptibility to a novel influenza virus. In the event of H7N9 virus, little is known about the level of pre-existing or cross-reactive immunity in humans. We measured haemagglutination inhibition (HAI) and neutralization antibody responses against the H7N9 virus in different age groups (children, 3–5 years old; adults, 18–59 years old, and old people, ≥ 60 years old) of pre- and post-seasonal influenza vaccination. No pre-existing immunity was detected in serum samples of all age groups. Also, no detectable cross-reactive antibodies against the H7N9 virus were present in all age groups after current seasonal vaccination (Table 1 and Supplementary Table 1). These data demonstrated that the human population is naive to the H7N9 virus, and current seasonal vaccination does not protect against H7N9 infection.

The antiviral drugs M2 ion channel blockers (amantadine and rimantadine) and neuraminidase inhibitors (oseltamivir, zanamivir, peramivir and laninamivir) are available at present for the prophylaxis and treatment of influenza virus infection. However, the H7N9 isolates contain one M2 blocker resistance-related mutation (Ser31Asn). Tested by our *in vitro* neuraminidase function assay, current H7N9 viruses are found to be sensitive to both oseltamivir and zanamivir (Supplementary Table 2). However, the antiviral resistance in A/H7N9 viruses should be a concern, as neuraminidase inhibitor-resistance mutants associated with a neuraminidase Arg292Lys substitution were recently observed in two patients infected with H7N9 and correlated with adverse clinical prognosis²⁵.

Although no efficient human-to-human transmission has occurred, biological features we demonstrated here, such as a dual receptor-binding preference and high growth ability, provide the H7N9 virus with higher transmissibility from avian to human. Together with the challenges in the available treatments for H7N9 infection caused by its clinical severity and emerging antiviral resistance, further unpredictable evolution and adaptation of the H7N9 virus and the lack of pre-existing immunity leave the human population at high risk. Thus, the threats of the H7N9 virus with pandemic potential should not be underestimated and intensive surveillance must be undertaken.

METHODS SUMMARY

Viruses and cells. All H7N9 avian influenza viruses were isolated and passaged in 9-day-old embryonated specific pathogen-free chicken eggs. All experiments with H7N9 viruses were performed in approved enhanced biosafety level 3 (BSL-3) containment laboratories. Human type II alveolar epithelial (A549) cells were maintained in DMEM, and Madin–Darby canine kidney (MDCK) and porcine kidney (PK15) cells were maintained in MEM, supplemented with 10% FBS, glutamine, HEPES, penicillin and streptomycin.

Direct binding assay with synthetic sialylglycopolymers. As described elsewhere⁵, serial dilutions of sialylglycopolymers in PBS were coated in 96-well flat-bottom polystyrene plates overnight at 4 °C. Thirty-two haemagglutination unit (HAU) live viruses per well, and oseltamivir carboxylate at a final concentration of 17.5 μ M were added, as described previously²⁶. After incubating with appropriate amounts of the primary antibodies, the secondary antibody was added, and bound antibody was detected using tetramethylbenzidine substrate solution.

Ex vivo infection of human lung and tracheal tissue by the H7N9 virus. Tissue fragments were infected with 2×10^7 TCID₅₀ (the tissue culture infectious dose affecting 50% of the cultures) A/Anhui/1/2013 virus for 24 h and 48 h, and the infectivity was assayed by immunohistochemistry staining and morphology examination. For replication tests, the tissue fragments were infected with 10^4 TCID₅₀ A/Anhui/1/2013 or A/Shanghai/1/2013 virus. The viral titres in the supernatants at 1, 24, 48 and 72 h.p.i. were measured in MDCK cells.

Cytokine and chemokine measurement. Sera concentrations of the chemokines and cytokines IP-10, MIG, MCP-1, MIP-1 β , IL-6, IL-8, IFN- α and IFN- γ were measured using cytometric bead-array assays according to the manufacturers' instructions.

Cross-reactive antibody assay. A total of 90 stored serum samples vaccinated with seasonal influenza vaccines between 2012 and 2013 were tested for antibody responses to the novel H7N9 subtype A/Anhui/1/2013 virus.

Full Methods and any associated references are available in the online version of the paper.

Received 1 May; accepted 14 June 2013.

Published online 3 July; corrected online 24 July 2013 (see full-text HTML version for details).

- Gao, R. *et al.* Human infection with a novel avian-origin influenza A (H7N9) virus. *N. Engl. J. Med.* **368**, 1888–1897 (2013).
- World Health Organization. Avian influenza A(H7N9) virus. http://www.who.int/influenza/human_animal_interface/influenza_h7n9/en/index.html (accessed, 31 May 2013).

3. Gao, H. N. *et al.* Clinical findings in 111 cases of influenza A (H7N9) virus infection. *N. Engl. J. Med.* **368**, 2277–2285 (2013).
4. Li, Q. *et al.* Preliminary report: epidemiology of the avian influenza A (H7N9) outbreak in China. *N. Engl. J. Med.* <http://dx.doi.org/10.1056/NEJMoa1304617> (24 April 2013).
5. Chandrasekaran, A. *et al.* Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. *Nature Biotechnol.* **26**, 107–113 (2008).
6. Connor, R. J., Kawaoka, Y., Webster, R. G. & Paulson, J. C. Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology* **205**, 17–23 (1994).
7. Matrosovich, M. *et al.* Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. *J. Virol.* **74**, 8502–8512 (2000).
8. Srinivasan, K., Raman, R., Jayaraman, A., Viswanathan, K. & Sasisekharan, R. Quantitative description of glycan-receptor binding of influenza A virus h7 hemagglutinin. *PLoS ONE* **8**, e49597 (2013).
9. Nidom, C. A. *et al.* Influenza A (H5N1) viruses from pigs, Indonesia. *Emerg. Infect. Dis.* **16**, 1515–1523 (2010).
10. Yang, H., Chen, L. M., Carney, P. J., Donis, R. O. & Stevens, J. Structures of receptor complexes of a North American H7N2 influenza hemagglutinin with a loop deletion in the receptor binding site. *PLoS Pathog.* **6**, e1001081 (2010).
11. Yamada, S. *et al.* Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. *Nature* **444**, 378–382 (2006).
12. Herfst, S. *et al.* Airborne transmission of influenza A/H5N1 virus between ferrets. *Science* **336**, 1534–1541 (2012).
13. Imai, M. *et al.* Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature* **486**, 420–428 (2012).
14. Walthers, T. *et al.* Glycomic analysis of human respiratory tract tissues and correlation with influenza virus infection. *PLoS Pathog.* **9**, e1003223 (2013).
15. Bateman, A. C. *et al.* Glycan analysis and influenza A virus infection of primary swine respiratory epithelial cells: the importance of NeuAc α 2–6 glycans. *J. Biol. Chem.* **285**, 34016–34026 (2010).
16. Pastva, A. M., Wright, J. R. & Williams, K. L. Immunomodulatory roles of surfactant proteins A and D: implications in lung disease. *Proc. Am. Thorac. Soc.* **4**, 252–257 (2007).
17. Gu, J. *et al.* H5N1 infection of the respiratory tract and beyond: a molecular pathology study. *Lancet* **370**, 1137–1145 (2007).
18. de Jong, M. D. *et al.* Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. *Nature Med.* **12**, 1203–1207 (2006).
19. Writing Committee of the Second World Health Organization Consultation on Clinical Aspects of Human Infection with Avian Influenza A (H5N1) Virus. Update on avian influenza A (H5N1) virus infection in humans. *N. Engl. J. Med.* **358**, 261–273 (2008).
20. Chen, Y. *et al.* Human infections with the emerging avian influenza A H7N9 virus from wet market poultry: clinical analysis and characterisation of viral genome. *Lancet* **381**, 1916–1925 (2013).
21. Tarazona, R. *et al.* Increased expression of NK cell markers on T lymphocytes in aging and chronic activation of the immune system reflects the accumulation of effector/senescent T cells. *Mech. Ageing Dev.* **121**, 77–88 (2001).
22. Freeman, C. M., Curtis, J. L. & Chensue, S. W. CC chemokine receptor 5 and CXC chemokine receptor 6 expression by lung CD8⁺ cells correlates with chronic obstructive pulmonary disease severity. *Am. J. Pathol.* **171**, 767–776 (2007).
23. Saetta, M. *et al.* Increased expression of the chemokine receptor CXCR3 and its ligand CXCL10 in peripheral airways of smokers with chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **165**, 1404–1409 (2002).
24. Shalekoff, S., Gray, G. E. & Tiemessen, C. T. Age-related changes in expression of CXCR4 and CCR5 on peripheral blood leukocytes from uninfected infants born to human immunodeficiency virus type 1-infected mothers. *Clin. Diagn. Lab. Immunol.* **11**, 229–234 (2004).
25. Hu, Y. *et al.* Association between adverse clinical outcome in human disease caused by novel influenza A H7N9 virus and sustained viral shedding and emergence of antiviral resistance. *Lancet* [http://dx.doi.org/10.1016/S0140-6736\(13\)61125-3](http://dx.doi.org/10.1016/S0140-6736(13)61125-3) (29 May 2013).
26. Childs, R. A. *et al.* Receptor-binding specificity of pandemic influenza A (H1N1) 2009 virus determined by carbohydrate microarray. *Nature Biotechnol.* **27**, 797–799 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements This study was supported by the National Basic Research Program (973) of China (2011CB504704, to Y.S.), Emergency Research Project on human infection with avian influenza H7N9 virus from the National Ministry of Science and Technology (no. KJYJ-2013-01-01 to Y.S.), and National Mega-projects for Infectious Diseases (2013ZX10004611 to J.Z.). We would like to acknowledge the Consortium for Functional Glycomics for providing the glycans for the receptor-binding test. We thank the US Centre for Disease Control and Prevention (CDC) for providing the strains A/Washington/1/2007(H3N2, WT), A/Texas/12/2007(H3N2, E119V), A/California/7/2009 (2009pdmH1N1, WT) and A/North Carolina/39/2009 (2009pdmH1N1, H275Y), and for the antiviral drug susceptibility test. We thank C. Li for critical comments of the manuscript and proofreading. We thank L. Huang, X. Xing, X. Zeng, Y. Zhu, L. Liu, S. Yang, Z. Li and T. Liu for technical support. We thank the provincial CDCs including Fujian, Sichuan, Guangdong, Hunan, Hubei and Jiangxi for providing sera specimens of patients with H5N1. The contents of this article are solely the responsibility of the authors and do not necessarily represent the views of the China CDC or other organizations.

Author Contributions Y.S., S.W., Z.F., J.H., W.Y., G.F.G., G.W., D.L. and Y.W. designed the experiments and discussed the results; J.Z., D.W., R.G., T.B., K.Q., Y.L., S.Z., J.G., J.D., L.D., Y.Z., H.W., X.L., J.L., L.L., X.Z., X.L., W.H., L.W., H.B., L.X., Y.C., C.X., S.W., Z.F., J.H., W.Y. and J.S. performed the experiments; B.Z., J.S., X.Q., Y.Z., Y.P., Y.Y., X.Z. and Y.S. collected and transferred samples and tissues. J.Z., D.W., R.G., L.Y. and W.Z. analysed data; J.Z., D.W., R.G., Y.L., W.Z. and Y.S. wrote the manuscript. J.Z., D.W., R.G. and B.Z. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.S. (yshu@cnic.org.cn).

METHODS

Cell culture. Human type II alveolar epithelial (A549), Madin-Darby canine kidney (MDCK) and porcine kidney (PK15) cells were obtained from the American Type Culture Collection. A549 cells were maintained in DMEM (Invitrogen) and MDCK and PK15 cells were in MEM (Invitrogen), supplemented with 10% FBS (Invitrogen), glutamine (2 mM; Invitrogen), HEPES (10 mM; Invitrogen), penicillin (100 U ml^{-1}), and streptomycin ($100 \mu\text{g ml}^{-1}$; Invitrogen).

Virus preparation, titration and infection. In view of the biosafety issues involved in handling H7N9 viruses, the experiments were conducted in approved enhanced biosafety level 3 (BSL-3) containment laboratories. The viruses were propagated in 9-day-old embryonated specific pathogen-free chicken eggs, by a first passage in the H7N9 strain. The allantoic fluid was collected at 48 h.p.i., aliquoted, and stored at -80°C until use. Infectious titres (TCID₅₀, the tissue culture infectious dose affecting 50% of the cultures) were performed in confluent MDCK cells and confirmed by haemagglutination assay.

Neuraminidase inhibitors. Oseltamivir carboxylate (4-*N*-acetyl-5-amino-3-(1-ethylpropoxy)-1-cyclohexane-1-carboxylic acid), the active metabolite of oseltamivir, was provided by Hoffmann-La Roche. Zanamivir (4-guanidino-Neu5Ac2en) was provided by GlaxoSmithKline.

Direct binding assay with synthetic sialylglycopolymers. Receptor specificity was investigated using biotinylated glycans including 3'-SLN, 3'-SLN-LN, 6'-SLN and 6'-SLN-LN (see legend to Fig. 1), which were obtained from the Consortium of Functional Glycomics through their resource request program. As described elsewhere⁵, serial dilutions of sialylglycopolymers in PBS were coated in 96-well flat-bottom polystyrene plates overnight at 4°C . Thirty-two haemagglutination unit (HAU) live viruses per well and oseltamivir carboxylate (Hoffmann-La Roche) at a final concentration of $17.5 \mu\text{M}$ were added, as described previously²⁶. After incubating with appropriate amounts of the primary antibodies, including a panel of rabbit sera against an Eurasian H7N7 virus isolated from the environment, A/California/07/09 (H1N1) and A/Anhui/01/2005 (H5N1) viruses, the secondary antibody was added (horseradish peroxidase (HRP)-conjugated goat anti-rabbit IgG; Santacruz Biotechnology), and bound antibody was detected using tetramethylbenzidine substrate solution (R&D Systems). All procedures were performed at 4°C to avoid the cleavage effects of neuraminidase on the binding.

Growth curves in vitro. A549, MDCK and PK15 cells were infected with A/Shanghai/1/2013, A/California/07/2009 or A/Anhui/1/2013 virus at a multiplicity of infection (MOI) of 0.001. One hour after inoculation at 37°C , at time point 0, the cells were washed once with PBS, and fresh infection medium containing 2 mg l^{-1} *N*-*p*-tosyl-L-phenylalaninechloromethyl ketone-treated (TPCK) trypsin (Sigma) was added. The supernatants were sampled at 12, 24, 48, 72 and 96 h.p.i. and the viral titres in these supernatants were determined by endpoint titration in MDCK cells as described²⁷.

Ex vivo infection of human lung and tracheal tissue by the H7N9 virus. Human tracheal and lung sections were obtained from one donor, and two additional lung fragments were obtained from another two donors of uninvolved portions of surgical resection samples from patients undergoing removal of lung tissue for possible carcinoma, and regions containing carcinoma or infection were excluded. All patients gave informed consent under a study approved by the Institutional Review Board of National Institute for Viral Disease Control and Prevention, China CDC. The tissues from each donor were cut into multiple 2–3-mm fragments and were infected with A/Anhui/1/2013 at a titre of 2×10^7 TCID₅₀ virus at 37°C for 24 h and 48 h and the infectivity was assayed by immunohistochemistry staining and morphology examination. As to the replication tests, the tissue fragments were infected with 10^4 TCID₅₀ A/Anhui/1/2013 or A/Shanghai/1/2013 virus. One hour after inoculation at 37°C , at time point 0, the tissues were washed three times with 2 ml PBS, and 500 μl fresh infection medium containing 2 mg l^{-1} TPCK trypsin was added. The supernatants were sampled at 1, 24, 48 and 72 h.p.i. and the viral titres in these supernatants were determined by endpoint titration in MDCK cells.

H&E staining and immunohistochemistry. Routine haematoxylin and eosin (H&E) staining was used to determine the histology of lung and trachea explants. To determine the viral infection in lung or bronchus tissues, immunohistochemical staining was performed on 4- μm -thick deparaffinized sections using monoclonal antibodies against the nucleoprotein of influenza A (VS2525, USCDC) by using a polymer-based colorimetric indirect immunalkaline phosphatase method (Thermo Scientific/Lab Vision). Sequential tissue sections incubated with normal serum pertinent to the primary antibody were used as negative controls. Paraffin-embedded sections were deparaffinized by two washes in xylene and were rehydrated through decreasing concentration of ethanol. After washing with PBS for 5 min at room temperature, sections were treated with $200 \mu\text{g ml}^{-1}$ proteinase K for 20 min. After blocking with UV blocking buffer (Thermo Scientific/Lab Vision), the sections were incubated with specific antibodies for 30 min at room temperature. Unbound antibody was removed by 0.5% PBS tween-20 (PBST)

before adding HRP-labelled polymer for 15 min at room temperature (Thermo Scientific/Lab Vision). After washing unbound labelled polymer, peroxidase staining in tissue sections was revealed by DAB solution (Thermo Scientific/Lab Vision). Double-stain immunohistochemistry was performed using peroxidase polymer-labelled antibodies (Thermo Scientific/Lab Vision/Dako) against cytokeratin AE1/AE3 (188132, Invitrogen) as the marker for epithelial cells, surfactant protein A as the marker for type II pneumocytes (ab51891, abcam) on the sections with the nucleoprotein staining, followed by the mouse anti-influenza A nucleoprotein antibody labelled with immunoalkaline phosphatase polymer. All staining steps were completed on an automatic staining system (DAKO, Denmark). After stopping the reaction in running water, sections were counterstained by a quick dip in a Mayer's haematoxylin solution. The sections were mounted with DPX and examined by light microscopy (ZEISS M2).

Cytokine and chemokine measurement. Sera concentrations of the chemokines/cytokines IP-10, MIG, MCP-1, MIP-1 β , IL-6, IL-8, IFN- α and IFN- γ from seven patients with H7N9 and ten close contacts were measured using cytometric bead-array assays (CBA; Becton Dickinson) according to the manufacturers' instructions. Fifty microlitres of tested sera or sera diluents in assay diluent was incubated with 50 μl capture beads suspension of each flex set kit per test at room temperature for 1.5 h, and then 50 μl phycoerythrin-conjugated detection reagent was added for another 1.5 h. After washing in 1 ml wash buffer, the samples were run and FACS data were collected using a FACSAria I flow cytometer and analysed using BD Cytometric Bead Array 1.4 software. The data together with those of 11 confirmed patients with H5N1 and 20 close contacts previously measured by CBA were analysed.

Cross-reactive antibody assay. We assessed the level of cross-reactive antibody to H7N9 virus in 90 subjects before and after they had been vaccinated with the 2012–2013 seasonal influenza vaccines. The compositions of the trivalent vaccine were A/Christchurch/16/2010 (H1N1), A/Victoria/361/2011 (H3N2) and B/Hubei-Wujiagang/158/2009. The HAI assay was according to standard protocols with 0.5% turkey red blood cells (TRBCs) and a microneutralization assay as the standard protocols were performed^{28,29}. Both HAI and microneutralization titres of 20 or more were defined as seropositive. We tested the samples for antibody responses to the novel H7N9 subtype viruses A/Anhui/1/2013. Before the HAI assay, serum samples were treated with a 1:4 (v/v) dilution of receptor-destroying enzyme (RDE; Denka Seiken) at 37°C for 18 h, followed by incubation at 56°C for 30 min. The serum samples for the microneutralization assay were inactivated at 56°C for 30 min. The acute and convalescent sera of two patients with H7N9 and animal sera against H5N1 or 2009 pdmH1N1 virus were included in the HAI and microneutralization tests as controls. For the HAI assay, 25- μl serial twofold sera diluents were incubated with standard virus of 4 HAU per 25 μl at room temperature for 30 min. Then, 50- μl standardized 0.5% TRBCs were added, mixed well and settled at room temperature for the appropriate 30 min. For the microneutralization assay, 50- μl serial twofold sera diluents were incubated with standard virus of 100 TCID₅₀ per 50 μl at 37°C , 5% CO₂ for 60 min. Then, 100 μl 1.5×10^5 MDCK cells per well was added for a further 18 h at 37°C , 5% CO₂. On day 2, the cells were fixed by 80% cold acetone and stained by anti-influenza A nucleoprotein monoclonal antibody pool (MAB8257 and MAB8258, Millipore) and peroxidase polymer-labelled goat anti-mice IgG (074-1802, Kirkegaard & Perry Laboratories). Bound antibody was detected using freshly prepared *o*-phenylenediamine dihydrochloride and citrate buffer (Sigma). The test dose of virus is acceptable if the back-titration is positive in 5–7 wells containing the lowest dilution of test virus.

Neuraminidase inhibition assay. Fluorimetric neuraminidase activity assays were done according to the manufacturers' instructions of the NA-FluorTM kit (Life Technologies). In brief, a standardized dose of viruses was incubated with various concentrations of oseltamivir carboxylate or zanamivir in assay buffer containing 2-(*N*-morpholino) ethanesulphonic acid and calcium chloride, for 45 min at 37°C . Then, substrate methylumbelliferyl-*N*-acetylneuraminic acid (MUNANA) was added as a fluorescent substrate to this mixture. After incubation for another hour at 37°C , stop solution was added to the mixture to stop the reaction. The fluorescence signal was measured at an excitation wavelength of 360 nm and an emission wavelength of 465 nm, and the 50% inhibitory concentrations (IC₅₀) of antiviral drugs were calculated according to the concentration required for reducing neuraminidase activity by 50% by analysis with GraphPad Prism software, version 5 (GraphPad Software).

Electron microscopic examination. The infected lung tissues with H7N9 virus were processed for transmission electron microscopy at 24 h post infection *ex vivo* as previously described³⁰. The AH1-infected lung tissues were cut into several 1-mm fragments and immersed in fixative (2% paraformaldehyde, 2.5% glutaraldehyde, 0.1% sodium cacodylate buffered, pH 7.2) for 1 h at 4°C . Tissues were then fixed in 1% osmium tetroxide (0.1% sodium cacodylate buffered, pH 7.2) for 1 h at 4°C . After several dehydrations and rehydrations, the blocks were infiltrated

with 1:1 (v/v) ethanol/epoxy resin (SPI-Chem low viscosity “Spurr” kits) for 1 h at room temperature following Spurr medium infiltration and embedded in BEEM embedding capsules with Spurr medium. Cured blocks were cut into ultrathin sections with a microtome (LKB Nova) and mounted on copper grids (150 meshes). The sections were stained with 1% uranyl acetate in 50% ethanol for 10 min and 0.2% lead citrate for 5 min, and then observed under transmission electron microscope (FEI Tecnai12) at 80 kV.

Statistical analysis. Statistical significance was determined using non-parametric tests and the GraphPad Prism 5 software package (GraphPad Software). $P < 0.05$ was deemed to indicate statistical significance.

27. Reed, L. J. & Muench, H. A simple method of estimating fifty percent endpoints. *Am. J. Hyg.* **27**, 493–497 (1938).
28. WHO. Collaborating Centers from Reference and Research on Influenza. *Concepts and Procedures for Laboratory-Based Influenza Surveillance* 17–35 (U.S. Department of Health and Human Services, 1982).
29. WHO Global Influenza Surveillance Network. Manual for the laboratory diagnosis and virological surveillance of influenza; http://whqlibdoc.who.int/publications/2011/9789241548090_eng.pdf.
30. Goldsmith, C. S. *et al.* Elucidation of Nipah virus morphogenesis and replication using ultrastructural and molecular approaches. *Virus Res.* **92**, 89–98 (2003).

Van der Waals heterostructures

A. K. Geim^{1,2} & I. V. Grigorieva¹

Research on graphene and other two-dimensional atomic crystals is intense and is likely to remain one of the leading topics in condensed matter physics and materials science for many years. Looking beyond this field, isolated atomic planes can also be reassembled into designer heterostructures made layer by layer in a precisely chosen sequence. The first, already remarkably complex, such heterostructures (often referred to as ‘van der Waals’) have recently been fabricated and investigated, revealing unusual properties and new phenomena. Here we review this emerging research area and identify possible future directions. With steady improvement in fabrication techniques and using graphene’s springboard, van der Waals heterostructures should develop into a large field of their own.

Graphene research has evolved into a vast field with approximately ten thousand papers now being published every year on a wide range of graphene-related topics. Each topic is covered by many reviews. It is probably fair to say that research on ‘simple graphene’ has already passed its zenith. Indeed, the focus has shifted from studying graphene itself to the use of the material in applications¹ and as a versatile platform for investigation of various phenomena. Nonetheless, the fundamental science of graphene remains far from being exhausted (especially in terms of many-body physics) and, as the quality of graphene devices continues to improve^{2–5}, more breakthroughs are expected, although at a slower pace.

Because most of the ‘low-hanging graphene fruits’ have already been harvested, researchers have now started paying more attention to other two-dimensional (2D) atomic crystals⁶ such as isolated monolayers and few-layer crystals of hexagonal boron nitride (hBN), molybdenum disulphide (MoS₂), other dichalcogenides and layered oxides. During the first five years of the graphene boom, there appeared only a few

experimental papers on 2D crystals other than graphene, whereas the last two years have already seen many reviews (for example, refs 7–11). This research promises to reach the same intensity as that on graphene, especially if the electronic quality of 2D crystals such as MoS₂ (refs 12, 13) can be improved by a factor of ten to a hundred.

In parallel with the efforts on graphene-like materials, another research field has recently emerged and has been gaining strength over the past two years. It deals with heterostructures and devices made by stacking different 2D crystals on top of each other. The basic principle is simple: take, for example, a monolayer, put it on top of another monolayer or few-layer crystal, add another 2D crystal and so on. The resulting stack represents an artificial material assembled in a chosen sequence—as in building with Lego—with blocks defined with one-atomic-plane precision (Fig. 1). Strong covalent bonds provide in-plane stability of 2D crystals, whereas relatively weak, van-der-Waals-like forces are sufficient to keep the stack together. The possibility of making multilayer van der Waals heterostructures has been demonstrated experimentally only

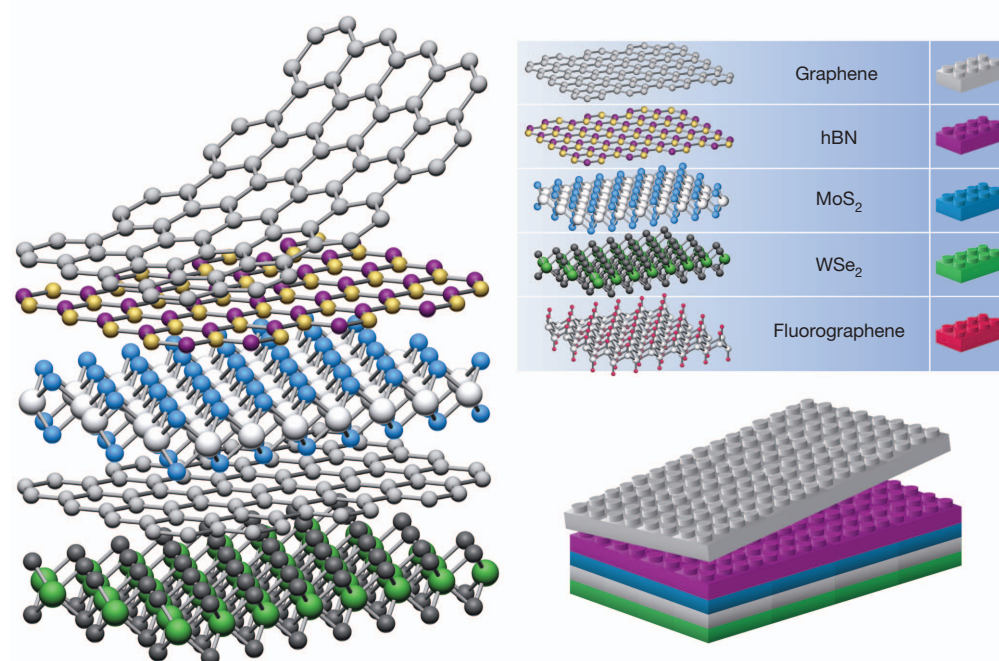


Figure 1 | Building van der Waals heterostructures. If one considers 2D crystals to be analogous to Lego blocks (right panel), the construction of a huge variety of layered structures becomes possible. Conceptually, this atomic-scale Lego resembles molecular beam epitaxy but employs different ‘construction’ rules and a distinct set of materials.

¹School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK. ²Centre for Mesoscience and Nanotechnology, University of Manchester, Manchester M13 9PL, UK.

recently^{14–19}. Most importantly, it turned out that in practice this atomic-scale Lego works exceptionally well, better than one could have imagined. How it works and why van der Waals heterostructures deserve attention is discussed in this Perspective.

Dreamscape

For any research subject, it is helpful to have a big idea, even if it is unlikely to be realized in its original form. In the case of graphene, its biggest ambition has been to grow into the new silicon, offering a lifeline for Moore's law¹. At the time of writing, the dreams of other 2D crystals are relatively more modest. They are often about offering alternative solutions to compensate for graphene's weaknesses^{7–11}. In contrast, van der Waals heterostructures do not lack ambition.

Imagine the following structure. Graphene is put on top of a dielectric crystal a few layers thick (for example, mica), and the sequence is repeated again and again. The resulting van der Waals crystal is superficially similar to superconducting copper oxides, with graphene playing the part of conductive CuO planes and the 2D high- κ dielectrics providing interplanar spacing. The critical temperature T_C of oxide superconductors depends on many materials parameters, including the CuO interlayer spacing^{20,21}. Careful tuning of these parameters allows T_C above 130 K. However, the standard growth techniques offer limited scope for varying the parameters, and progress has stalled. What if we mimic layered superconductors by using atomic-scale Lego? Bismuth strontium calcium copper oxide superconductors (BSCCO) can be disassembled into individual atomically thin planes⁶. Their reassembly with some intelligently guessed differences seems worth a try, especially when the mechanism of high- T_C superconductivity remains unknown. Moreover, graphene seems to be a natural choice of 2D component in the search for new layered superconductors. Indeed, intercalated graphite exhibits a respectable T_C above 10 K (ref. 22) and can be viewed as a stack of heavily doped graphene planes with an increased interlayer distance. What if a dielectric plane of, for example, BSCCO or hBN is added in between the planes of intercalated graphite? Such artificial materials engineered with one-atomic-plane accuracy would have been science fiction a few years ago but are within the grasp of today's technology.

We have used the above example for its straightforward appeal and because high- T_C superconductivity in doped graphene has been widely proposed (see refs 23–26). Yet van der Waals heterostructures bring to mind not one but many similar speculative ideas. Another example is a room-temperature excitonic superfluidity suggested for two graphene layers separated by an ultrathin dielectric^{27,28}. On the scale of such dramatic perspectives, it may sound fairly modest that van der Waals heterostructures also offer a helping hand in graphene's efforts to go beyond silicon^{1,15}. Of course, one can think of many arguments why these or similar ideas can fail. In blue-sky research, even the most plausible scenarios often do. However, big dreams are essential to keep us trying and to serve as Ariadne's thread when exploring new topics. Having said that, it is equally important not to get lost on the way by checking grand ideas against contemporary reality.

Layered reality check

Before explaining how to make van der Waals heterostructures, it is instructive to review the existing library of 2D crystals, those individual components that can be used in the assembly. In principle, there exist hundreds of layered materials that cleave easily, and one can naturally think of using, for example, the same Scotch-tape cleaving technique to isolate their atomic planes⁶. Unfortunately, this is not straightforward. One must remember that (1) melting temperature decreases with decreasing the thickness of thin films and (2) most materials survive our ambient conditions only by natural passivation of their surfaces²⁹. Monolayers have two surfaces and no bulk in between, which presents the extreme case of surface science. On the other hand, neither of the procedures developed for isolating 2D crystals can currently be carried out in high vacuum or at low temperature, which are typical surface-science requirements. And graphene monolayers are notably more reactive than even

graphene bilayers^{30,31}. In short, many 2D crystals imaginable in theory are unlikely to survive in reality because they would corrode, decompose, segregate and so forth²⁹.

The high thermal and chemical stability of a three-dimensional (3D) crystal is essential before one can even contemplate the possibility of its 2D counterpart. Graphite has both, allowing graphene to exist under ambient conditions. The same is valid for other stable 2D crystals in Fig. 2. Nonetheless, the surface of MoS₂ starts oxidizing in moist air below 100 °C (ref. 32). Even graphene would not survive if our rooms were twice as hot, at 600 K (ref. 30). Take, for example, GaSe, TaS₂ or Bi₂Se₃. These materials are stable in bulk but, once cleaved down to a few layers, corrode. An illuminating case is silicene^{33,34}. This 2D silicon can be grown epitaxially and investigated in high vacuum. However, it is not expected to survive isolation from its parent substrate or exposure to air. With contemporary technologies, the silicene transistors envisaged in the literature cannot be made. These examples are given to say that, because of the poor stability of atomically thin films compared to their 3D counterparts, the library of 2D crystals should be relatively limited.

Another important consideration is interfacial contamination. Adsorbates such as water, hydrocarbons and so on cover every surface, unless it is prepared under extreme surface-science conditions. Graphene is densely covered with hydrocarbons, even after annealing in the high vacuum of a transmission electron microscope. It takes considerable effort to find clean patches several nanometres in size¹⁶. (Note that this contamination is highly mobile and usually remains unnoticeable for scanning probe microscopy.) If isolated 2D crystals are stacked together, the surface contamination becomes trapped in between layers. Therefore, van der Waals heterostructures should generally be expected to become 'layer cakes' glued by contamination rather than the neat crystals imagined in Fig. 1. This scenario would remove much of the appeal from van der Waals heterostructures because it is difficult to control the manufacture and reproducibility of 'layer cakes'. Fortunately, it turns out that contamination can clean itself off the interfaces^{16,19,35}, as further discussed below.

Two-dimensional family values

At the time of writing, we can be certain of the existence of more than a dozen different 2D crystals under ambient conditions. First of all, these are monolayers of graphite, hBN and MoS₂, which have been studied extensively. It is probably not coincidental that these materials are widely used as solid lubricants, which requires high thermal and chemical stability. There are also 2D tungsten disulphide (WS₂), tungsten diselenide (WSe₂) and molybdenum diselenide (MoSe₂), which are chemically, structurally and electronically similar to MoS₂. Despite little research having been done so far on the latter monolayers, it is safe to add them to the 2D library, too (Fig. 2).

Among the above 2D crystals, graphene is an unequivocal champion, exhibiting the highest mechanical strength and crystal and electronic quality. It is likely to be the most common component in future van der Waals heterostructures and devices. Latest developments on graphene include micrometre-scale ballistic transport at room temperature (refs 3, 36) and low-temperature carrier mobilities of $\mu \approx 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ in suspended devices^{4,5}. The runner-up is 2D hBN, or 'white graphene'. Its rise started when bulk hBN crystals^{37,38} were shown to be an exceptional substrate for graphene, allowing a tenfold increase in its electronic quality². This advance attracted immediate attention and, shortly after, few-layer crystals and monolayers of hBN were used as gate dielectrics^{14,39} and tunnel barriers (2D hBN can sustain biases up to about 0.8 V nm^{-1} and be free from pinholes)^{40,41}.

Monolayers of MoS₂ were studied earlier^{6,42}, including the demonstration of the electric field effect, but they received little attention until devices with switching on/off ratios of $>10^8$ and room-temperature $\mu \approx 100 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ were reported¹². Although these mobilities are much lower than in graphene, they are still remarkably high compared with thin-film semiconductors. The large on/off ratios are due to a sizeable bandgap in MoS₂. It is direct in a monolayer (about 1.8 eV), whereas bilayer and few-layer MoS₂ are indirect bandgap semiconductors^{43,44}.

Graphene family	Graphene	hBN 'white graphene'	BCN	Fluorographene	Graphene oxide
2D chalcogenides	MoS ₂ , WS ₂ , MoSe ₂ , WSe ₂		Semiconducting dichalcogenides: MoTe ₂ , WTe ₂ , ZrS ₂ , ZrSe ₂ and so on	Metallic dichalcogenides: NbSe ₂ , NbS ₂ , TaS ₂ , TiS ₂ , NiSe ₂ and so on	
				Layered semiconductors: GaSe, GaTe, InSe, Bi ₂ Se ₃ and so on	
2D oxides	Micas, BSCCO	MoO ₃ , WO ₃	Perovskite-type: LaNb ₂ O ₇ , (Ca,Sr) ₂ Nb ₃ O ₁₀ , Bi ₄ Ti ₃ O ₁₂ , Ca ₂ Ta ₂ TiO ₁₀ and so on		Hydroxides: Ni(OH) ₂ , Eu(OH) ₂ and so on
	Layered Cu oxides	TiO ₂ , MnO ₂ , V ₂ O ₅ , TaO ₃ , RuO ₂ and so on			Others

Figure 2 | Current 2D library. Monolayers proved to be stable under ambient conditions (room temperature in air) are shaded blue; those probably stable in air are shaded green; and those unstable in air but that may be stable in inert atmosphere are shaded pink. Grey shading indicates 3D compounds that have been successfully exfoliated down to monolayers, as is clear from atomic force microscopy, for example, but for which there is little further information. The

data given are summarized from refs 6–11, 42 and 55. We note that, after intercalation and exfoliation, the oxides and hydroxides may exhibit stoichiometry different from their 3D parents (for example, TiO₂ exfoliates into a stoichiometric monolayer of Ti_{0.87}O₂; ref. 8). 'Others' indicates that many other 2D crystals—including borides, carbides, nitrides and so on—have probably been^{7–11} or can be isolated. BCN, boron carbon nitride.

Semiconductors with a direct gap are of special interest for use in optics and optoelectronics. Further interest in monolayer MoS₂ is due to the broken centrosymmetry that allows efficient spin and valley polarization by optical pumping^{45,46}. This research is stimulated by the availability of large molybdenite crystals from several mining sources. The absence of such supply is probably the reason why 2D WS₂, WSe₂ and MoSe₂ attract relatively little attention despite the fact that Raman and transport studies have revealed their electronic structures and quality to be similar to that of MoS₂ (refs 47–49). The differences between these dichalcogenides worth noting are the stronger spin–orbit coupling in the W compounds and the lower stability of the Se compounds.

There have been reports on exfoliation of many other layered chalcogenides down to a monolayer (Fig. 2). However, we have chosen only to 'pencil' them into the 2D library because little is known about their stability, let alone their optical and transport properties. In some cases, it is even unclear whether the observed flakes and suspensions are indeed 2D counterparts of the parent crystals or present different chemical entities after exposure to air or liquids. In our experience, monolayers of metallic dichalcogenides are unstable in air (T. Georgiou, R. R. Nair, A. Mishchenko, K. S. Novoselov & A.K.G., manuscript in preparation).

Another group of 2D crystals are numerous oxides including monolayers of TiO₂, MoO₃, WO₃, mica and perovskite-like crystals such as BSCCO and Sr₂Nb₃O₁₀ (for reviews, see refs 7–11). As oxides, these crystals are less susceptible to air but they tend to lose oxygen and may react with minority chemicals (for example, water and hydrogen). Similarly to other atomically thin crystals, properties of 2D oxides are expected to differ from those of their parents owing to quantum confinement. Indeed, monolayer oxides often have lower dielectric constants and larger bandgaps than their 3D counterparts⁸ and can exhibit charge density waves⁶. Unfortunately, information about oxide monolayers is limited mostly to their observation with atomic force and electron microscopes. The heavy artillery of physics and nanotechnology such as the electric field effect, Raman and optical spectroscopy, tunnelling, and so on have not yet been applied to isolated 2D oxides. We should also mention 2D hydroxides that can be exfoliated down to a monolayer but even less is known about them⁷.

Finally, there are several graphene derivatives (Fig. 2). One of them is fluorographene, a stable stoichiometric wide-gap insulator, which can be obtained by fluorination of graphene⁵⁰. Unfortunately, only crystals with poor electronic quality have been reported so far. Graphane (fully hydrogenated graphene) gradually loses its hydrogen³¹ and is unlikely to be useful for making heterostructures. Nonetheless, we note that hydrogenated or other derivatives can sometimes be more stable than 2D crystals themselves⁵¹. Finally, let us mention graphene oxide⁵² and monolayers of boron carbon nitride^{53,54}, which although non-stoichiometric, can also be considered for designing van der Waals heterostructures.

Rules of survival

As interest in graphene-like crystals rapidly grows^{7–11}, the search for new 2D candidates is expected to intensify, too. In this regard, the following rule of thumb can be helpful. First, 3D materials with melting temperature over 1,000 °C have the best chances of having 2D counterparts stable at room temperature. Second, 3D parents must be chemically inert and exhibit no decomposed surface layer in air or an alternative environment where exfoliation takes place. Third, insulating and semiconducting 2D crystals are more likely to be stable than metallic ones, owing to the generally higher reactivity of metals. In all cases, visual evaluation and Raman spectroscopy are helpful to provide a rapid test for the absence of corrosion and the presence of essential signatures indicating a similarity to the parent crystal. However, the ultimate proof lies with electrical measurements of either in-plane transport for conducting 2D crystals or out-of-plane tunnelling through insulating ones to check for their homogeneity and the absence of pinholes.

As a further step towards expanding the 2D library, one can perform isolation and encapsulation in an inert atmosphere. Many metallic 2D dichalcogenides may then remain stable at room temperature, as their stability in solvents seems to indicate^{55,56}. This approach can also lead to higher electronic quality for present favourites such as graphene and 2D MoS₂. Exfoliation-encapsulation at low temperature (such as in liquid nitrogen) is in principle possible but for the moment too difficult to contemplate for practical use. Lastly, monolayers may exist without a layered 3D parent (examples are silicene and monolayers of Y₂O₃ and ZnO)^{57,58}. If the monolayers are sufficiently stable, the substrate can be etched away, as demonstrated for graphene grown on metal foils^{59–61}. This can provide access to 2D crystals without 3D layered analogues in nature.

Lego on atomic scale

It is no longer adventurous to imagine the automated, roll-to-roll assembly⁶¹ of van der Waals heterostructures using sheets of epitaxially grown 2D crystals^{59–63}. However, concerted efforts towards such assembly are expected only when a particular heterostructure proves to be worthy of attention, as happened in the case of graphene on hBN⁶⁴. For scouting which area to focus on, manual assembly is likely to remain the favourite approach. It offers high throughput and relatively easy changes in layer sequences. Likewise, individual 2D compounds will continue to be obtained by the Scotch-tape technique, which has so far provided crystallites of unmatched quality. Nevertheless, we expect the increasingly frequent use of epitaxially grown graphene, 2D hBN, 2D MoS₂, and so on for making proof-of-concept van der Waals heterostructures.

At the time of writing, only a few groups have reported van der Waals heterostructures made from more than two atomically thin crystals, and only graphene and few-layer hBN, MoS₂ and WS₂ were used for this

assembly^{14–19,65,66}. A typical stacking procedure starts by isolating micro-metre-sized 2D crystals on top of a thin transparent film (for example, polymer). The resulting 2D crystal provides one brick for the Lego wall in Fig. 1 and can now be put face down onto a chosen target. The supporting film is then removed or dissolved. More 2D crystals are produced, and the transfer is repeated again and again, until a desired stack is assembled. Conceptually, this is simple and requires only basic facilities such as a good optical microscope. In practice, the fabrication technique takes months to master. In addition to the standard clean-room procedures (cleaning, dissolving, resist spinning and so on), it is necessary to position different 2D crystals over each other with micro-metre accuracy. This is done under the microscope by using micromanipulators. The crystals must be put in soft contact without rubbing and, ideally no liquid or polymer should be allowed in contact with cleaved surfaces to minimize contamination. Thermal annealing in an inert atmosphere can often be helpful after adding each new layer. For transport measurements, 2D crystals are plasma etched into, for example, Hall bars with contacts evaporated as the final step.

Despite the dozens of steps involved, sophisticated multilayer structures can now be produced within a matter of days. Figure 3 shows two such examples. One is a van der Waals superlattice made from six alternating bilayers of graphene and hBN. This is the largest number of 2D crystals in a van der Waals heterostructure reported so far¹⁶. The most challenging design has probably been double-layer graphene devices¹⁸ such as those shown in Fig. 3b, c. We emphasize that interfaces in these heterostructures are found to be clean and atomically sharp^{16,19}, without the contaminating ‘goo’ that always covers 2D crystals even in high

vacuum (see the ‘Layered reality check’ section). The reason for the clean interfaces is the van der Waals forces that attract adjacent crystals and effectively squeeze out trapped contaminants or force them into micro-metre-sized ‘bubbles’¹⁶. This allows 10- μm -scale devices that are effectively free from contamination. We also note that atomically sharp interfaces are in practice impossible to achieve by other techniques, including molecular beam epitaxy, because of island growth.

Little evolutionary steps

Although the availability of various isolated 2D crystals had been recognized⁶, practical steps towards their van der Waals assembly were taken only after 2010. An important stimulus was the demonstration that hBN could serve as a high-quality substrate for graphene² (many other substrates, including pyrolytic hBN, were unsuccessfully tried before^{35,67}). This led to rapid development of transfer procedures. The next logical step was encapsulation, where thin hBN crystals served not only as the substrate but also as a protective cover for graphene³. Encapsulation has proved its worth by enabling devices with consistently high quality that do not deteriorate under ambient conditions. The most commonly achieved mobility for graphene on hBN is $\mu \approx 100,000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ but up to $500,000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ can be reached at low temperature. Such high quality (as indicated by high mobility) can be witnessed directly as negative bend resistance and magnetic focusing^{3,36} (Fig. 4a). These ballistic effects persist up to room temperature. The encapsulation also results in high spatial uniformity so that capacitors over $100 \mu\text{m}^2$ in size exhibit quantum oscillations in magnetic fields B as low as 0.2 T (ref. 68).

The next evolutionary step has been ‘vertical’ devices in which few-layer-thick crystals of hBN, MoS_2 or WS_2 are used as tunnel barriers with graphene serving as one or both electrodes^{15,19}. These devices require three to four transfers, but no plasma etching, to define their geometry. Although sensitive to charge inhomogeneity, vertical devices usually do not pose critical demands on μ . The tunnelling heterostructures allow the demonstration of a new kind of electronic device: field-effect tunnelling transistors¹⁵. In these, the tunnel current is controlled by changes in the electrode’s Fermi energy, which can be varied by gate voltage by as much as about 0.5 eV owing to the low density of states in monolayer graphene. An increase in the Fermi energy effectively lowers the tunnel barrier, even if no bias is applied¹⁵. This is in contrast to the standard Fowler–Nordheim mechanism, which is based on tilting the top of the tunnel barrier by applied bias. The van der Waals tunnelling devices exhibit an on/off switching ratio of over 10^6 at room temperature^{15,19,69}.

A higher level of complexity is presented by the graphene–hBN superlattice shown in Fig. 3a. It proves the concept that thin films of new 3D materials consisting of dozens of atomic layers are in principle possible by reassembly, as discussed in the ‘Dreamscape’ section. In the case of Fig. 3a, hBN bilayers serve as spacers whereas bilayer graphene (rather than its monolayer) was chosen to facilitate intercalation to reach a high density of states. Further efforts in making and investigating such multilayer structures are expected, given the interest generated by a large amount of literature on possible collective phenomena in graphene-based systems^{23–28,70–72}.

The double-layer devices in Fig. 3 represent the state of the art for van der Waals heterostructures. They were designed to probe in-plane transport in the regime of the strongest possible electron–electron interaction between electrically isolated 2D systems^{18,73}. The separation of the graphene layers can be as small as three hBN layers (about 1 nm)¹⁸ but this still provides a sufficiently high potential barrier to suppress electron tunnelling. The layers continue to ‘feel’ each other strongly through Coulomb interactions. The 1-nm separation is much smaller than the in-plane distance between charge carriers in graphene, which is typically around 10 nm and nominally diverges near the neutrality point. This makes the interlayer separation the smallest spatial parameter in the problem. Therefore, the two electronic liquids in double-layer graphene effectively nest within the same plane, but can still be tuned and measured separately.

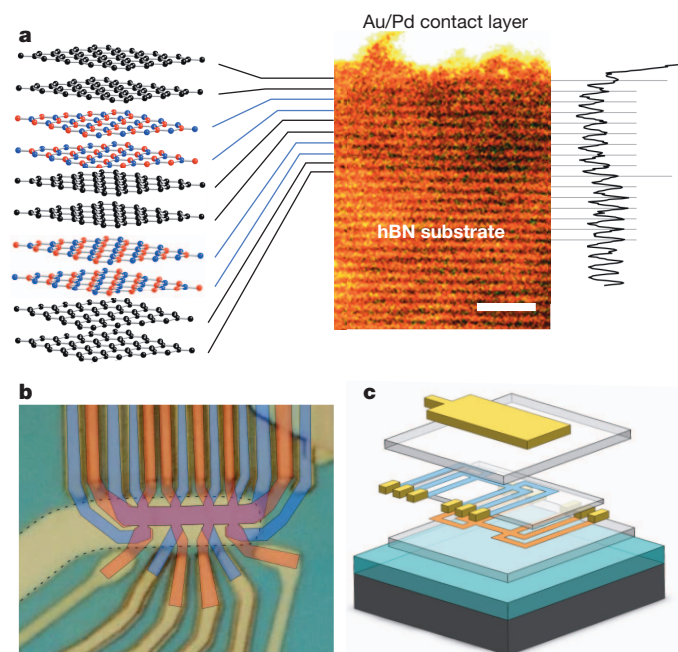


Figure 3 | State-of-the-art van der Waals structures and devices.

a, Graphene–hBN superlattice consisting of six stacked bilayers. On the right its cross-section and intensity profile as seen by scanning transmission electron microscopy are shown; on the left is a schematic view of the layer sequence. The topmost hBN bilayer is not visible, being merged with the metallic contact. Scale bar, 2 nm. (Adapted from ref. 16.) **b**, **c**, Double-layer graphene heterostructures¹⁸. An optical image of a working device (**b**) and its schematics in matching colours (**c**). Two graphene Hall bars are accurately aligned, separated by a trilayer hBN crystal and encapsulated between relatively thick hBN crystals (hBN is shown in **c** as semitransparent slabs). The entire heterostructure is placed on top of an oxidized Si wafer (SiO_2 is in turquoise). The colours in **b** indicate the top (blue) and bottom (orange) Hall bars and their overlapping region (violet). The graphene areas are invisible in the final device image because of the top Au gate outlined by dashes. The scale is given by the width of the Hall bars, 1.5 μm .

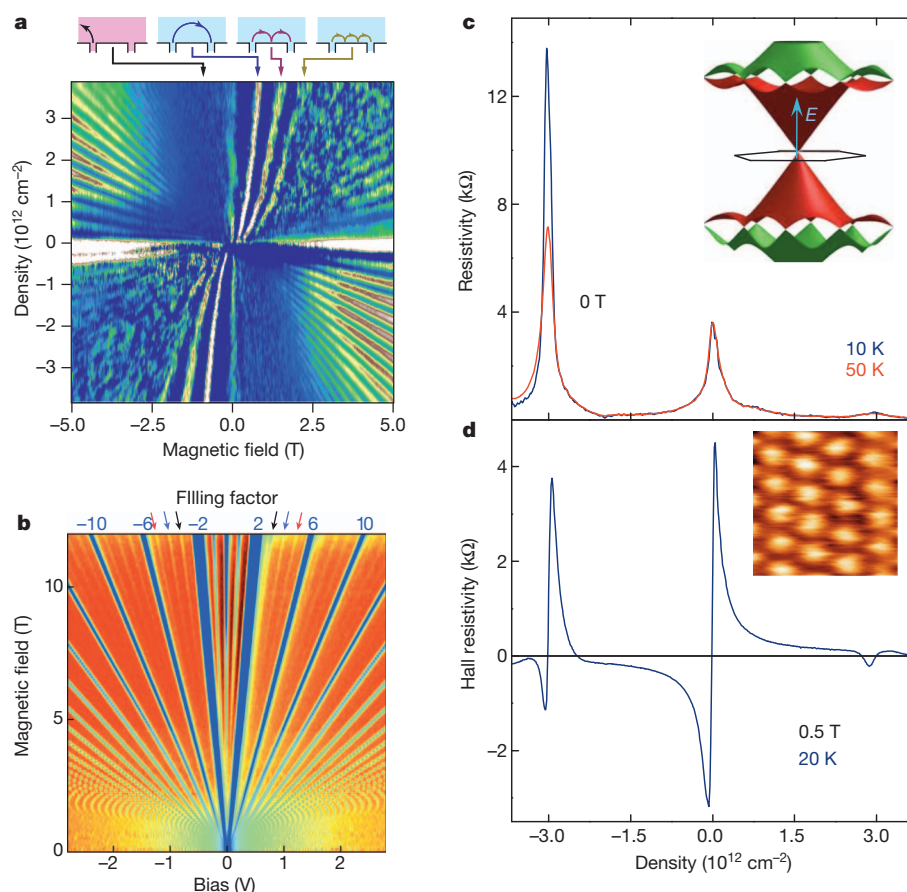


Figure 4 | Early harvest in van der Waals fields. **a**, Magnetic focusing in graphene on hBN. Pronounced resonances are observed if the size of a cyclotron orbit becomes commensurate with the distance between narrow graphene leads used as an injector and a detector (adapted from ref. 36). Bright colours show maxima in conductivity as a function of carrier density (positive and negative densities correspond to electrons and holes, respectively) and B . The upper panel illustrates the corresponding trajectories. **b**, Quantum capacitance of encapsulated graphene as a function of gate voltage and B . In this device, spin and valley degeneracies are lifted above 10 T. (Adapted from ref. 68.) **c**, **d**, Importance of crystallographic alignment. The standard Dirac-like spectrum is strongly reconstructed for graphene on hBN, and new Dirac cones appear in both valence and conduction bands (inset in **c**). This leads to pronounced peaks in resistivity (**c**) and the Hall effect changes sign (**d**). The inset in **d** shows the moiré patterns that lead to the spectral reconstruction. (Adapted from ref. 76.)

The ambition behind the double-layer heterostructures is to address many possible collective states, including Wigner crystallization, excitonic superfluidity, itinerant magnetism and so on. Such many-body phenomena have so far been the domain of low-temperature physics. In particular, Bose condensation of interlayer excitons has been reported at sub-Kelvin temperatures in double-layer semiconductor heterostructures, and Coulomb drag has been used as a way of detecting the condensate^{74,75}. The hope was that double-layer graphene would allow a similar superfluid state but at much higher temperature^{27,28} because the Coulomb energy of interlayer interaction can exceed 0.1 eV. The measurements¹⁸ revealed many interesting and some unexpected features but no sign of superfluidity at zero magnetic field B so far. By analogy with the semiconductor heterostructures^{74,75}, the regime of quantizing B remains the most promising for the observation of coherent electronic states, but this has not been investigated yet.

Another important step in the sophistication of van der Waals heterostructures was to include crystallographic alignment, which can be done with an accuracy of less than 1° (ref. 76). Although the interaction between stacked 2D crystals is relatively weak (around 10 meV \AA^{-2} ; ref. 77), electron orbitals still extend out of the plane and affect charge carriers in an adjacent 2D crystal. This influence results in moiré patterns that depend on the rotation angle between joining crystals and their lattice mismatch^{78–80} (Fig. 4d). In the case of graphene on hBN, a periodic potential created by the hBN substrate results in additional Dirac cones at high electron and hole densities^{76,81} (Fig. 4c). The superlattice effects are remarkably strong, such that the Hall effect changes its sign (Fig. 4d) and the new Dirac cones exhibit their own sets of Landau levels^{76,81}. This is the best proof so far that interfacial contamination can be negligible and also shows that the electronic spectra of van der Waals heterostructures can be tuned by using moiré potentials. As another example of such bandgap engineering, one can consider the alignment of monolayers that have close lattice constants (for example, MoS_2 and WSe_2). The resulting van der Waals crystal is expected to exhibit optical and electronic properties distinct from its components⁸².

Many other types of van der Waals structure and device are expected to be demonstrated soon, initially using only a small number of 2D crystals. Among obvious objectives are various proximity effects. To this end, 2D crystals can be put on top of atomically flat crystals exhibiting magnetism, ferroelectricity, spin–orbit coupling and so on. For example, graphene encapsulated in WS_2 is likely to exhibit an induced spin–orbit interaction that should affect transport properties.

Handicraft on industrial scale

The growing interest in van der Waals heterostructures is not limited to new physics and materials science. There is also a massive potential for applications. Here we avoid early speculations because interest in van der Waals heterostructures is already justified if one considers them as a way of accelerating the development of the myriad applications offered by graphene itself. The recently demonstrated new graphene-based device architectures^{1,15,83} provide straightforward examples.

Any industrial application will require a scalable approach to van der Waals assembly. To this end, significant efforts have been reported to grow graphene, 2D hBN and 2D MoS_2 epitaxially on top of each other^{10,84–88}. However, it is a daunting task to find the right conditions for so-called van der Waals epitaxy⁸⁹ because the weak interlayer interaction generally favours island growth rather than that of continuous monolayers. Another scalable approach is layer-by-layer deposition from 2D-crystal suspensions by using Langmuir–Blodgett or similar techniques^{8,90}. One can also mix suspensions of different 2D crystals and then make layer-by-layer laminates, relying on self-organizational assembly (flocculation)¹¹. Unfortunately, micrometre-sized crystallites in suspensions cannot provide large continuous layers, and this would limit possible applications for such van der Waals laminates. At present, they are being considered for use as designer ultrathin dielectrics⁸, selectively permeable membranes⁹¹ and composite materials⁹².

At the time of writing, the most feasible approach to industrial-scale production of van der Waals heterostructures seems to involve growing

individual monolayers on catalytic substrates, then isolating and stacking these 2D sheets. This route has already been proved to be scalable^{61,64}. If a particular heterostructure attracts sufficient interest in terms of applications, it seems inevitable that its production can be scaled up by trying a variety of available approaches.

Long live graphene

After many years of intensive effort, graphene research should logically have reached a mature stage. However, the possibility of combining graphene with other 2D crystals has expanded this field dramatically, well beyond simple graphene or 2D MoS₂. The interest in van der Waals heterostructures is growing as quickly as interest in graphene did a few years ago. As the technology of making van der Waals heterostructures moves from its humble beginnings, increasingly sophisticated devices and materials should become available to more and more research groups. This is likely to cause a snowball effect because, with so many 2D crystals, sequences and parameters to consider, the choice of possible van der Waals structures is limited only by our imagination. Even with the 2D components that have been shown to be stable, it will take time and effort to explore the huge parameter space. The decades of research on semiconductor heterostructures and devices may serve as a guide to judge the probable longevity of research on van der Waals materials, beyond simple graphene.

Received 1 April; accepted 12 June 2013.

- Novoselov, K. S. *et al.* A roadmap for graphene. *Nature* **490**, 192–200 (2012).
- Dean, C. R. *et al.* Boron nitride substrates for high-quality graphene electronics. *Nature Nanotechnol.* **5**, 722–726 (2010).
- This paper attracted attention to hBN as a substrate and initiated the development of transfer techniques essential for the van der Waals reassembly.**
- Mayorov, A. S. *et al.* Micrometer-scale ballistic transport in encapsulated graphene at room temperature. *Nano Lett.* **11**, 2396–2399 (2011).
- Mayorov, A. S. *et al.* How close can one approach the Dirac point in graphene experimentally? *Nano Lett.* **12**, 4629–4634 (2012).
- Bao, W. *et al.* Evidence for a spontaneous gapped state in ultraclean bilayer graphene. *Proc. Natl Acad. Sci. USA* **109**, 10802–10805 (2012).
- Novoselov, K. S. *et al.* Two-dimensional atomic crystals. *Proc. Natl Acad. Sci. USA* **102**, 10451–10453 (2005).
- This was the first paper to demonstrate the electric field effect and study electron transport in 2D crystals other than graphene.**
- Mas-Ballester, R., Gómez-Navarro, C., Gómez-Herrero, J. & Zamora, F. 2D materials: to graphene and beyond. *Nanoscale* **3**, 20–30 (2011).
- Osada, M. & Sasaki, T. Two-dimensional dielectric nanosheets: novel nanoelectronics from nanocrystal building blocks. *Adv. Mater.* **24**, 210–228 (2012).
- Wang, Q. H., Kalantar-Zadeh, K., Kis, A., Coleman, J. N. & Strano, M. S. Electronics and optoelectronics of two-dimensional transition metal dichalcogenides. *Nature Nanotechnol.* **7**, 699–712 (2012).
- Xu, M., Lian, T., Shi, M. & Chen, H. Graphene-like two-dimensional materials. *Chem. Rev.* **113**, 3766–3798 (2013).
- We recommend this review for initial acquaintance with 2D materials other than graphene.**
- Butler, S. Z. *et al.* Progress, challenges, and opportunities in two-dimensional materials beyond graphene. *ACS Nano* **7**, 2898–2926 (2013).
- Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V. & Kis, A. Single-layer MoS₂ transistors. *Nature Nanotechnol.* **6**, 147–150 (2011).
- The paper attracted critical attention to electron transport in MoS₂ monolayers.**
- Fuhrer, M. S. & Hone, J. Measurement of mobility in dual-gated MoS₂ transistors. *Nature Nanotechnol.* **8**, 146–147 (2013).
- Ponomarenko, L. A. *et al.* Tunable metal-insulator transition in double-layer graphene heterostructures. *Nature Phys.* **7**, 958–961 (2011).
- This is the first demonstration of multilayer van der Waals heterostructures, beyond using hBN, mica and so on as a substrate.**
- Britnell, L. *et al.* Field-effect tunneling transistor based on vertical graphene heterostructures. *Science* **335**, 947–950 (2012).
- Haigh, S. J. *et al.* Cross-sectional imaging of individual layers and buried interfaces of graphene-based heterostructures and superlattices. *Nature Mater.* **11**, 764–767 (2012).
- The paper proves the concept of complex heterostructures, including manually assembled van der Waals superlattices, and shows that their interfaces can be atomically sharp and clean.**
- Dean, C. R. *et al.* Graphene based heterostructures. *Solid State Commun.* **152**, 1275–1282 (2012).
- Gorbachev, R. V. *et al.* Strong Coulomb drag and broken symmetry in double-layer graphene. *Nature Phys.* **8**, 896–901 (2012).
- Georgiou, T. *et al.* Vertical field-effect transistor based on graphene–WS₂ heterostructures for flexible and transparent electronics. *Nature Nanotechnol.* **8**, 100–103 (2013).
- Kastner, M. A., Birgeneau, R. J., Shirane, G. & Endoh, Y. Magnetic, transport, and optical properties of monolayer copper oxides. *Rev. Mod. Phys.* **70**, 897–928 (1998).
- Orenstein, J. & Millis, A. J. Advances in the physics of high-temperature superconductivity. *Science* **288**, 468–474 (2000).
- Weller, T. E., Ellerby, M., Saxena, S. S., Smith, R. P. & Skipper, N. T. Superconductivity in the intercalated graphite compounds C₆H₁₈Yb and C₆Ca. *Nature Phys.* **1**, 39–41 (2005).
- Profeta, G., Calandra, M. & Mauri, F. Phonon-mediated superconductivity in graphene by lithium deposition. *Nature Phys.* **8**, 131–134 (2012).
- Nandkishore, R., Levitov, L. S. & Chubukov, A. V. Chiral superconductivity from repulsive interactions in doped graphene. *Nature Phys.* **8**, 158–163 (2012).
- Savini, G., Ferrari, A. C. & Giustino, F. First-principles prediction of doped graphene as a high-temperature electron-phonon superconductor. *Phys. Rev. Lett.* **105**, 037002 (2010).
- Guinea, F. & Uchoa, B. Odd-momentum pairing and superconductivity in vertical graphene heterostructures. *Phys. Rev. B* **86**, 134521 (2012).
- Min, H., Bistrizter, R., Su, J. J. & MacDonald, A. H. Room-temperature superfluidity in graphene bilayers. *Phys. Rev. B* **78**, 121401 (2008).
- Perali, A., Neilson, D. & Hamilton, A. R. High-temperature superfluidity in double-bilayer graphene. *Phys. Rev. Lett.* **110**, 146803 (2013).
- Geim, A. K. Random walk to graphene. *Rev. Mod. Phys.* **83**, 851–862 (2011).
- Liu, L. *et al.* Graphene oxidation: thickness-dependent etching and strong chemical doping. *Nano Lett.* **8**, 1965–1970 (2008).
- Elias, D. C. *et al.* Control of graphene's properties by reversible hydrogenation: evidence for graphane. *Science* **323**, 610–613 (2009).
- Ross, S. & Sussman, A. Surface oxidation of molybdenum disulfide. *J. Phys. Chem.* **59**, 889–892 (1955).
- Vogt, P. *et al.* Silicene: compelling experimental evidence for graphenelike two-dimensional silicon. *Phys. Rev. Lett.* **108**, 155501 (2012).
- Fleurence, A. *et al.* Experimental evidence for epitaxial silicene on diboride thin films. *Phys. Rev. Lett.* **108**, 245501 (2012).
- Lui, C. H., Liu, L., Mak, K. F., Flynn, G. W. & Heinz, T. F. Ultraflat graphene. *Nature* **462**, 339–341 (2009).
- Taychatanapat, T., Watanabe, K., Taniguchi, T. & Jarillo-Herrero, P. Electrically tunable transverse magnetic focusing in graphene. *Nature Phys.* **9**, 225–229 (2013).
- Watanabe, K., Taniguchi, T. & Kanda, H. Direct-bandgap properties and evidence for ultraviolet lasing of hexagonal boron nitride single crystal. *Nature Mater.* **3**, 404–409 (2004).
- Zomer, P. J., Dash, S. P., Tombros, N. & van Wees, B. J. A transfer technique for high mobility graphene devices on commercially available hexagonal boron nitride. *Appl. Phys. Lett.* **99**, 232104 (2011).
- Meric, I. *et al.* Graphene field-effect transistors based on boron nitride gate dielectrics. *Tech. Digest Int. Electron Devices Meet. 2010 IEEE Int.* **10**, 556–559, doi:10.1109/IEDM.2010.5703419 (2010).
- Lee, G. H. *et al.* Electron tunneling through atomically flat and ultrathin hexagonal boron nitride. *Appl. Phys. Lett.* **99**, 243114 (2011).
- Britnell, L. *et al.* Electron tunneling through ultrathin boron nitride crystalline barriers. *Nano Lett.* **12**, 1707–1710 (2012).
- Gordon, R. A., Yang, D., Crozier, E. D., Jiang, D. T. & Frindt, R. F. Structures of exfoliated single layers of WS₂, MoS₂, and MoSe₂ in aqueous suspension. *Phys. Rev. B* **65**, 125407 (2002).
- Splendiani, A. *et al.* Emerging photoluminescence in monolayer MoS₂. *Nano Lett.* **10**, 1271–1275 (2010).
- Mak, K. F., Lee, C., Hone, J., Shan, J. & Heinz, T. F. Atomically thin MoS₂: a new direct-gap semiconductor. *Phys. Rev. Lett.* **105**, 136805 (2010).
- Zeng, H., Dai, J., Yao, W., Xiao, D. & Cui, X. Valley polarization in MoS₂ monolayers by optical pumping. *Nature Nanotechnol.* **7**, 490–493 (2012).
- Mak, K. F., He, K., Shan, J. & Heinz, T. F. Control of valley polarization in monolayer MoS₂ by optical helicity. *Nature Nanotechnol.* **7**, 494–498 (2012).
- Fang, H. *et al.* High-performance single layered WSe₂ p-FETs with chemically doped contacts. *Nano Lett.* **12**, 3788–3792 (2012).
- Zhao, W. *et al.* Evolution of electronic structure in atomically thin sheets of WS₂ and WSe₂. *ACS Nano* **7**, 791–797 (2013).
- Tonndorf, P. *et al.* Photoluminescence emission and Raman response of monolayer MoS₂, MoSe₂, and WSe₂. *Opt. Express* **21**, 4908–4916 (2013).
- Nair, R. R. *et al.* Fluorographene: a two-dimensional counterpart of Teflon. *Small* **6**, 2877–2884 (2010).
- Bianco, E. *et al.* Stability and exfoliation of germanene: a germanium graphene analogue. *ACS Nano* **7**, 4414–4421 (2013).
- Park, S. & Ruoff, R. S. Chemical methods for the production of graphenes. *Nature Nanotechnol.* **4**, 217–224 (2009).
- Jin, Z., Yao, J., Kittrell, C. & Tour, J. M. Large-scale growth and characterizations of nitrogen-doped monolayer graphene sheets. *ACS Nano* **5**, 4112–4117 (2011).
- Ci, L. *et al.* Atomic layers of hybridized boron nitride and graphene domains. *Nature Mater.* **9**, 430–435 (2010).
- Gamble, F. R. *et al.* Intercalation complexes of Lewis bases and layered sulfides: a large class of new superconductors. *Science* **174**, 493–497 (1971).
- Coleman, J. N. *et al.* Two-dimensional nanosheets produced by liquid exfoliation of layered materials. *Science* **331**, 568–571 (2011).
- Addou, R., Dahal, A. & Batzill, M. Growth of a two-dimensional dielectric monolayer on quasi-freestanding graphene. *Nature Nanotechnol.* **8**, 41–45 (2013).
- Tusche, C., Meyerheim, H. L. & Kirschner, J. Observation of depolarized ZnO(0001) monolayers: formation of unreconstructed planar sheets. *Phys. Rev. Lett.* **99**, 026102 (2007).

59. Reina, A. *et al.* Large area, few-layer graphene films on arbitrary substrates by chemical vapor deposition. *Nano Lett.* **9**, 30–35 (2009).
60. Li, X. *et al.* Large-area synthesis of high-quality and uniform graphene films on copper foils. *Science* **324**, 1312–1314 (2009).
61. Bae, S. *et al.* Roll-to-roll production of 30-inch graphene films for transparent electrodes. *Nature Nanotechnol.* **5**, 574–578 (2010).
62. Song, L. *et al.* Large scale growth and characterization of atomic hexagonal boron nitride layers. *Nano Lett.* **10**, 3209–3215 (2010).
63. Kim, K. K. *et al.* Synthesis of monolayer hexagonal boron nitride on Cu foil using chemical vapor deposition. *Nano Lett.* **12**, 161–166 (2012).
64. Bresnahan, M. S. *et al.* Integration of hexagonal boron nitride with quasi-freestanding epitaxial graphene: toward wafer-scale, high-performance devices. *ACS Nano* **6**, 5234–5241 (2012).
65. Bertolazzi, S., Krasnozhan, D. & Kis, A. Nonvolatile memory cells based on MoS₂/graphene heterostructures. *ACS Nano* **7**, 3246–3252 (2013).
66. Hunt, B. *et al.* Massive Dirac fermions and Hofstadter butterfly in a van der Waals heterostructure. *Science* **340**, 1427–1430 (2013).
67. Ponomarenko, L. A. *et al.* Effect of a high-kappa environment on charge carrier mobility in graphene. *Phys. Rev. Lett.* **102**, 206603 (2009).
68. Yu, G. L. *et al.* Interaction phenomena in graphene seen through quantum capacitance. *Proc. Natl Acad. Sci. USA* **110**, 3282–3286 (2013).
69. Yang, H. *et al.* Graphene barristor, a triode device with a gate-controlled Schottky barrier. *Science* **336**, 1140–1143 (2012).
70. Kotov, V. N., Pereira, V. M., Castro Neto, A. H. & Guinea, F. Electron–electron interactions in graphene: current status and perspectives. *Rev. Mod. Phys.* **84**, 1067–1125 (2012).
71. Das Sarma, S., Adam, S., Hwang, E. H. & Rossi, E. Electronic transport in two dimensional graphene. *Rev. Mod. Phys.* **83**, 407–470 (2011).
72. McChesney, J. L. *et al.* Extended van Hove singularity and superconducting instability in doped graphene. *Phys. Rev. Lett.* **104**, 136803 (2010).
73. Tutuc, E. & Kim, S. Magnetotransport and Coulomb drag in graphene double layers. *Solid State Commun.* **15**, 1283–1288 (2012).
74. Eisenstein, J. P. & MacDonald, A. H. Bose–Einstein condensation of excitons in bilayer electron systems. *Nature* **432**, 691–694 (2004).
75. Nandi, D., Finck, A. D. K., Eisenstein, J. P., Pfeiffer, L. N. & West, K. W. Exciton condensation and perfect Coulomb drag. *Nature* **488**, 481–484 (2012).
76. Ponomarenko, L. A. *et al.* Cloning of Dirac fermions in graphene superlattices. *Nature* **497**, 594–597 (2013).
77. Björkman, T., Gulans, A., Krasheninnikov, A. V. & Nieminen, R. M. van der Waals bonding in layered compounds from advanced density-functional first-principles calculations. *Phys. Rev. Lett.* **108**, 235502 (2012).
78. Li, G. *et al.* Observation of Van Hove singularities in twisted graphene layers. *Nature Phys.* **6**, 109–113 (2010).
79. Decker, R. *et al.* Local electronic properties of graphene on a BN substrate via scanning tunneling microscopy. *Nano Lett.* **11**, 2291–2295 (2011).
80. Yankowitz, M. *et al.* Emergence of superlattice Dirac points in graphene on hexagonal boron nitride. *Nature Phys.* **8**, 382–386 (2012).
81. Dean, C. R. *et al.* Hofstadter's butterfly and fractal quantum Hall effect in moiré superlattices. *Nature* **497**, 598–602 (2013).
82. Kośmider, K. & Fernández-Rossier, J. Electronic properties of the MoS₂–WS₂ heterojunction. *Phys. Rev. B* **87**, 075451 (2013).
83. Kim, K., Choi, J. Y., Kim, T., Cho, S. H. & Chung, H. J. A role for graphene in silicon-based semiconductor devices. *Nature* **479**, 338–344 (2011).
84. Tanaka, T., Ito, A., Tajiima, A., Rokuta, E. & Oshima, C. Heteroepitaxial film of monolayer graphene/monolayer h-BN on Ni(111). *Surf. Rev. Lett.* **10**, 721–726 (2003).
85. Yan, Z. *et al.* Growth of bilayer graphene on insulating substrates. *ACS Nano* **5**, 8187–8192 (2011).
86. Liu, Z. *et al.* Direct growth of graphene/hexagonal boron nitride stacked layers. *Nano Lett.* **11**, 2032–2037 (2011).
87. Garcia, J. M. *et al.* Graphene growth on h-BN by molecular beam epitaxy. *Solid State Commun.* **152**, 975–978 (2012).
88. Shi, Y. *et al.* Van der Waals epitaxy of MoS₂ layers using graphene as growth templates. *Nano Lett.* **12**, 2784–2791 (2012).
89. Koma, A. Van der Waals epitaxy—a new epitaxial growth method for a highly lattice-mismatched system. *Thin Solid Films* **216**, 72–76 (1992).
90. Ariga, K., Ji, Q., Hill, J. P., Bando, Y. & Aono, M. Forming nanomaterials as layered functional structures toward materials nanoarchitectonics. *NPG Asia Mater.* **4**, e17, doi:10.1038/am.2012.30 (2012).
91. Nair, R. R., Wu, H. A., Jayaram, P. N., Grigorieva, I. V. & Geim, A. K. Unimpeded permeation of water through helium-leak-tight graphene-based membranes. *Science* **335**, 442–444 (2012).
92. Young, R. J., Kinloch, I. A., Gong, L. & Novoselov, K. S. The mechanics of graphene nanocomposites: a review. *Compos. Sci. Technol.* **72**, 1459–1476 (2012).

Acknowledgements We thank all participants of the Friday Graphene Seminar in Manchester for discussions, and R. Gorbachev and J. Chapman for help with the figures. This work was supported by the Royal Society, the European Research Council, the Körber Foundation, the Office of Naval Research and the Air Force Office of Scientific Research.

Author Contributions A.K.G. wrote a draft that was scrutinized and improved by both authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence should be addressed to I.V.G. (irina.grigorieva@man.ac.uk).

SAR11 viruses and defensive host strains

ARISING FROM Y. Zhao *et al.* *Nature* **494**, 357–360 (2013)

The recent findings of abundant SAR11 viruses by Zhao *et al.*¹ are intriguing, and add new insight into the on-going discussion of why SAR11 bacteria are highly successful in the pelagic ocean. On the basis of high SAR11 virus abundance, Zhao *et al.*¹ claim that SAR11 bacteria are competition specialists. Alternatively, we show here how their findings could be consistent with a dominance of defensive SAR11 strains. Considering their high abundance, understanding why SAR11 bacteria are so successful has important implications for the study of the pelagic ecosystem². There is a Reply to this Brief Communication Arising by Giovannoni, S., Temperton, B. & Zhao, Y. *Nature* **499**, <http://dx.doi.org/10.1038/nature12388> (2013).

Zhao *et al.*¹ interpret the high number of SAR11 viruses as an indication that SAR11 bacteria are competition strategists, not very successful in viral defence. This would match earlier findings of their small cell size and streamlined genome, which were previously interpreted as competitive traits of SAR11 (ref. 3). They explain the co-existence of abundant SAR11 bacteria and viruses with the ‘Red Queen Hypothesis’ (RQH) and state an apparent conflict with the ‘Kill-the-Winner’ model (KtW).

The prevailing idea that SAR11 bacteria are either competition strategists^{1,3} or defence strategists⁴ becomes less absolute once strains are introduced in conceptual models. Assuming that growth rate and viral defence of host strains are inversely related through a cost of resistance (COR)⁵, several theoretical possibilities are conceivable. In the ‘clustered scenario’ (Fig. 1a), strains of species are clustered along the growth-rate axis. This is predicted if between-species differences in maximum growth rates are large and species-specific CORs decrease with decreasing maximum growth rates. Given their high abundance, this scenario suggests that SAR11 bacteria are represented by the slow-growing, dark-shaded species (Fig. 1a). However, because the KtW with COR predicts inverse rank-abundance curves of hosts and their associated viruses (Fig. 1a, b), this is in conflict with high abundances of the recently discovered SAR11 viruses. Alternatively, if maximum growth rates and CORs are comparable between species, strains from different species would be interspersed along the growth-rate axis⁵ (‘interspersed scenario’, Fig. 1b). Species are then no longer either defensive or competitive. This scenario is consistent with high abundances of SAR11 bacteria and viruses and a dominance of defensive SAR11 strains if SAR11 is represented by the dark-shaded species (Fig. 1b).

A key, albeit often misunderstood, prediction of the KtW is that viral abundances are positively correlated with differences in host growth rates, not with differences in abundance of co-existing hosts⁶. Hence, if SAR11 strains cover a wide range of growth rates, highly abundant SAR11 viruses are predicted. Most of these viruses would key in on the fastest growing strains, keeping these ‘winner strains’ at low abundance. Slow-growing strains consequently dominate. Interestingly, Malmstrom *et al.*⁷ found growth rates of different SAR11 bacteria indeed covering the entire growth-rate spectrum of other prokaryotes in the Sargasso Sea. Owing to these findings, the previously observed slow growth rates of abundant SAR11 bacteria⁸ and the predicted high SAR11 virus abundance in the ‘interspersed scenario’, we favour this scenario. Furthermore, a dominance of defence strategists is consistent with perturbation experiments on viral control, and seems to hold even in oligotrophic environments^{9,10}. A more definite answer to SAR11’s dominant life strategies, however, requires further investigations into strain-specific growth rates and specificity of the SAR11 viruses discovered.

Zhao *et al.*¹ state a conflict in the RQH and KtW. We rather perceive the two concepts as models describing different aspects of microbial

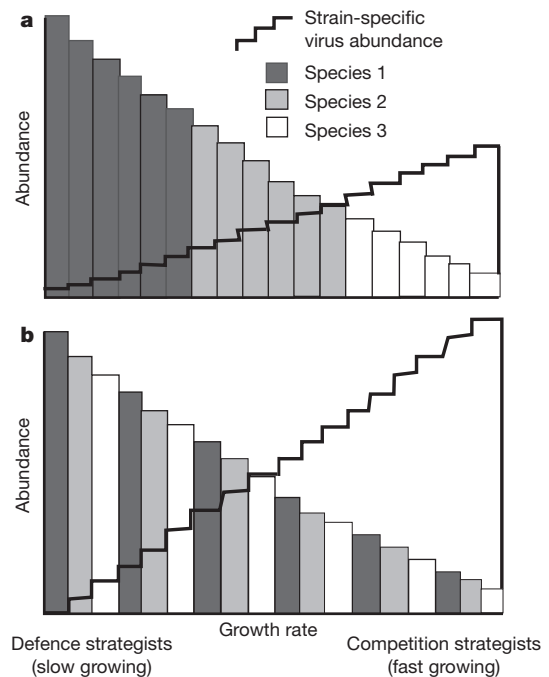


Figure 1 | Clustered versus interspersed distribution of bacterial strains along the growth-rate axis. **a**, ‘Clustered scenario’: strains of species are clustered in a narrow window along the growth-rate axis. In this case, the KtW predicts modest virus abundances within the species, and highest virus abundance for the fast-growing, white species (competition strategist). Given their high abundance, SAR11 bacteria are represented by the dark-shaded species in this scenario. **b**, ‘Interspersed scenario’: strains of all species are dispersed over the entire growth-rate axis, SAR11 bacteria being represented by the dark-shaded species. The KtW predicts highest total and within-species virus abundances in this scenario. Note the inverse rank-abundance curves of hosts and their associated viruses in the KtW with trade-off between host growth rate and viral defence. In this discussion, strain-specific one-to-one infections are assumed, although the model gives similar abundance predictions assuming a more general infection pattern with cross-infection⁵. Figure adapted from Våge *et al.*⁵.

interactions. The RQH is based on evolutionary dynamics occurring at evolutionary time scales, whereas KtW focuses on steady-state situations that may occur at ecological timescales. The two processes are not mutually exclusive. Rather, KtW explains diversity beyond spatial and temporal variability. Also, abundance fluctuations of strains may in fact be controlled by KtW processes^{9,10}.

Selina Våge¹, Julia E. Storesund¹ & T. Frede Thingstad¹

¹Department of Biology, University of Bergen, 5020 Bergen, Norway.
email: Selina.Vage@bio.uib.no

Received 18 March; accepted 12 June 2013.

1. Zhao, Y. *et al.* Abundant SAR11 viruses in the ocean. *Nature* **494**, 357–360 (2013).
2. Kirchman, D. L. Microbial oceanography: killers of the winners. *Nature* **494**, 320–321 (2013).
3. Grote, J. *et al.* Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *mBio* **3**, e00252–12 (2012).
4. Suttle, C. A. Marine viruses — major players in the global ecosystem. *Nature Rev. Microbiol.* **5**, 801–812 (2007).

5. Våge, S., Storesund, J. E. & Thingstad, T. F. Adding a cost of resistance description extends the ability of virus-host model to explain observed patterns in structure and function of pelagic microbial communities. *Environ. Microbiol.* **15**, 1842–1852 (2013).
 6. Thingstad, T. F. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol. Oceanogr.* **45**, 1320–1328 (2000).
 7. Malmstrom, R. R., Kiene, R. P., Cottrell, M. T. & Kirchman, D. L. Contribution of SAR11 bacteria to dissolved dimethylsulfoniopropionate and amino acid uptake in the North Atlantic Ocean. *Appl. Environ. Microbiol.* **70**, 4129–4135 (2004).
 8. Teira, E., Martinez-Garcia, S., Lønborg, C. & Alvarez-Salgado, X. A. Growth rates of different phylogenetic bacterioplankton groups in a coastal upwelling system. *Environ. Microbiol. Rep.* **1**, 545–554 (2009).
 9. Bouvier, T. & del Giorgio, P. A. Key role of selective viral-induced mortality in determining marine bacterial community composition. *Environ. Microbiol.* **9**, 287–297 (2007).
 10. Rodriguez-Brito, B. *et al.* Viral and microbial community dynamics in four aquatic environments. *ISME J.* **4**, 739–751 (2010).
- Author Contributions** S.V., J.E.S. and T.F.T. were responsible for planning, discussing and editing the comment. T.F.T. and S.V. developed the model, and S.V. was responsible for writing and illustrating the comment.
- Competing Financial Interests** Declared none.
- doi:10.1038/nature12387

Giovannoni *et al.* reply

REPLYING TO S. Våge, J. E. Storesund & T. F. Thingstad *Nature* **499**, <http://dx.doi.org/10.1038/nature12387> (2013)

In the debate about top-down versus bottom-up control, most would argue that both are important. The most probable scenario is that in the absence of high rates of transport, the equilibrium between these competing processes determines bacterial abundance. In their Brief Communication Arising, Våge and co-workers¹ respond to our Letter², arguing the case for defence specialism. ‘Defence specialism’ is not synonymous with ‘defence’. Because of the prevalence of viruses, it is probable that most bacteria have acquired defences against predation. ‘Defence specialism’ refers to success by defence, in the sense that a cell might allocate a large proportion of its resources to defence and thereby become successful.

Våge and co-workers¹ make a convincing argument that the defence specialism hypothesis is compatible with observations we report², if a range of growth rates and levels of defence specialism are allowed. Their argument postulates an inverse relationship between defence and growth rate and makes the reasonable assumption that there is a cost of resistance (COR). We agree that the scenario suggested by Våge and co-workers¹ is plausible and in our paper² report a number of genome features that might be defensive, and are variable properties of SAR11 cells³. However, an earlier investigation measuring COR in *Synechococcus* showed that moderate to large CORs were not universal among resistant strains and that large CORs tended to be associated with specific viral strains⁴. The COR of putative viral-defence-associated adaptations in SAR11 remains untested.

The central question addressed in our letter² is whether viral defence explains SAR11 success, that is, are they defence specialists. Although defence specialism remains a viable idea, it would need to be shown that an organism spends a large proportion of its resources defending, and is unusual (at one end of a distribution) in that regard. Våge *et al.*¹ assume that variation in ³H-labelled amino-acid-uptake rates among SAR11 cells, shown in Malmstrom *et al.*⁵, is evidence for differential growth rates, which they equate with degrees of defence specialism within a community of SAR11 quasi-species. However, Malmstrom *et al.*⁵ interpreted their data as evidence for variation in SAR11 cell ‘activity’, which can be uncoupled from growth. Because natural SAR11 are mixtures of ecotypes optimized to different oceanic conditions⁶ at any one time, even if one assumes a universal degree of defence specialism, the ecotype model predicts a range of SAR11 activities within a single water sample. Second, in such measurements, the apparent activities of cells will probably vary by at least a factor of two over a growth cycle. Therefore, we submit that although one interpretation of the data in Malmstrom could indeed be variability in defence specialism, other sources of variation are expected.

There is strong evidence that members of the SAR11 clade have an unusual metabolic strategy for resource competition^{7–9} and metagenomic, transcriptomic and proteomic data support the conclusion

that a significant part of the proteome is dedicated to nutrient uptake, both in culture and *in situ*^{10,11}. Until more robust evidence of defence specialism in SAR11 is available, we would argue that in the face of demonstrated pelagiphage abundance, a more cautious interpretation is that its success is more likely due to superior resource competition.

In our Letter we propose that recombination offers a positive feedback mechanism that enables winners to retain their dominance, and cite evidence for high recombination rates in SAR11 populations. We propose calling this the ‘King of the Mountain’ (KoM) hypothesis, after a popular children’s game. There is no conflict between the mechanism of KoM and that of the Kill-the-Winner (KtW) hypothesis, or the Red Queen Hypothesis. Rather, these ideas describe fundamental, compatible classes of interaction that are likely to affect population dynamics in conjunction with resource competition and growth.

This is not a trivial matter, because of the implications for geochemical cycles. Both the KtW and KoM hypotheses predict increased rates of immunity to viruses, but by very different mechanisms. In KtW the cost comes at the expense of competitiveness for resource acquisition, whereas KoM is a positive feedback loop that rewards superior resource competition, increasing contributions to geochemical cycles.

As the Brief Communication Arising largely relates to the physiology of SAR11, we have limited the authorship of this Reply to the current authors with whom the greatest expertise in this field resides.

Stephen Giovannoni¹, Ben Temperton¹ & Yanlin Zhao¹

¹Department of Microbiology, Oregon State University, Corvallis, Oregon 97331, USA.

email: steve.giovannoni@oregonstate.edu

1. Våge, S., Storesund, J. E. & Thingstad, T. F. SAR11 viruses and defensive host strains. *Nature* **499**, <http://dx.doi.org/10.1038/nature12387> (2013).
2. Zhao, Y. *et al.* Abundant SAR11 viruses in the ocean. *Nature* **494**, 357–360 (2013).
3. Grote, J. *et al.* Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *mBio* **3**, e00252–12 (2012).
4. Lennon, J. T. *et al.* Is there a cost of virus resistance in marine cyanobacteria? *ISME J.* **1**, 300–312 (2007).
5. Malmstrom, R. R., Kiene, R. P., Cottrell, M. T. & Kirchman, D. L. Contribution of SAR11 bacteria to dissolved dimethylsulfoniopropionate and amino acid uptake in the North Atlantic ocean. *Appl. Environ. Microbiol.* **70**, 4129–4135 (2004).
6. Vergin, K. L. *et al.* High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. *ISME J.* **7**, 1322–1332 (2013).
7. Tripp, H. J. *et al.* SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature* **452**, 741–744 (2008).
8. Tripp, H. J. *et al.* Unique glycine-activated riboswitch linked to glycine-serine auxotrophy in SAR11. *Environ. Microbiol.* **11**, 230–238 (2009).

BRIEF COMMUNICATIONS ARISING

9. Sun, J. *et al.* One carbon metabolism in SAR11 pelagic marine bacteria. *PLoS ONE* **6**, e23973 (2011).
10. Sowell, S. M. *et al.* Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *ISME J.* **3**, 93–105 (2008).
11. Sowell, S. M. *et al.* Proteomic Analysis of Stationary Phase in the Marine Bacterium 'Candidatus Pelagibacter ubique'. *Appl. Environ. Microbiol.* **74**, 4091–4100 (2008).
doi:10.1038/nature12388

Structure of the human glucagon class B G-protein-coupled receptor

Fai Yiu Siu¹, Min He², Chris de Graaf³, Gye Won Han¹, Dehua Yang², Zhiyun Zhang², Caihong Zhou², Qingping Xu⁴, Daniel Wacker¹, Jeremiah S. Joseph¹, Wei Liu¹, Jesper Lau⁵, Vadim Cherezov¹, Vsevolod Katritch¹, Ming-Wei Wang² & Raymond C. Stevens¹

Binding of the glucagon peptide to the glucagon receptor (GCGR) triggers the release of glucose from the liver during fasting; thus GCGR plays an important role in glucose homeostasis. Here we report the crystal structure of the seven transmembrane helical domain of human GCGR at 3.4 Å resolution, complemented by extensive site-specific mutagenesis, and a hybrid model of glucagon bound to GCGR to understand the molecular recognition of the receptor for its native ligand. Beyond the shared seven transmembrane fold, the GCGR transmembrane domain deviates from class A G-protein-coupled receptors with a large ligand-binding pocket and the first transmembrane helix having a 'stalk' region that extends three alpha-helical turns above the plane of the membrane. The stalk positions the extracellular domain (~12 kilodaltons) relative to the membrane to form the glucagon-binding site that captures the peptide and facilitates the insertion of glucagon's amino terminus into the seven transmembrane domain.

The glucagon receptor (GCGR) is one of the 15 members of the secretin-like (class B) family of G-protein-coupled receptors (GPCRs)¹ in humans. GCGR is activated by the 29 amino acid hormonal peptide glucagon (Supplementary Fig. 1a), and is a potential drug target for type 2 diabetes². During fasting, the pancreas dispatches glucagon to activate GCGR in the liver causing the release of glucose into the blood². Despite less than 15% protein sequence homology between class A (rhodopsin-like) and class B GPCRs, many of these receptors presumably share a seven transmembrane (7TM) helical domain and similar signal transduction mechanisms¹. Although the structure-function understanding of the class A family of GPCRs has been greatly advanced during the last few years³, a detailed understanding of class B GPCRs has lagged due to the lack of a 7TM domain structure for these receptors.

Secretin-like GPCRs contain a globular N-terminal extracellular domain (ECD) defined by three conserved disulphide bonds^{4,5} and a 7TM domain. They are activated by hormonal peptides that bind to both the ECD and the 7TM domain⁴. Structural details of soluble ECDs, including the ECD of GCGR⁶, and their role in the selective recognition of peptide hormones' carboxy termini have been revealed for several class B receptors^{5,7,8}. Although models of class B 7TM domains and ligand binding have been proposed based on site-directed mutagenesis^{9–11}, photo-crosslinking^{12–14}, and structure-based virtual screening studies¹⁵, the accuracy of such modelling has been hampered by the low sequence homology between class A and class B GPCRs.

Crystal structure of GCGR 7TM domain

The 7TM domain of human GCGR was fused to the thermally stabilized *E. coli* apocytochrome *b*₅₆₂RIL (ref. 16) (referred to as BRIL) at residue 123, and the C terminus of GCGR was truncated at residue 432 (Supplementary Fig. 2). This crystallized GCGR construct with BRIL containing a truncated ECD (ΔECD) and C terminus (ΔC) (BRIL-GCGR(ΔECD/ΔC), Supplementary Fig. 3) has the same binding affinity for the antagonist ligand NNC0640 (Supplementary Fig. 1b) as the full-length wild-type GCGR (Supplementary Table 1), indicating

that the conformation of the 7TM domain of BRIL-GCGR(ΔECD/ΔC) is similar to wild-type GCGR. The structure of the BRIL-GCGR(ΔECD/ΔC) was determined at 3.4 Å resolution (Methods and Supplementary Table 2). Although GCGR was crystallized in the presence of NNC0640, convincing electron density for NNC0640 was not observed. As expected, GCGR adopts a 7TM fold (Fig. 1), with the BRIL fusion protein folded on top of the receptor and mediating most of the crystal contacts (Supplementary Fig. 4).

Despite the lack of protein sequence conservation, comparison of the GCGR 7TM structure with 15 known class A GPCR structures solved in inactive form shows that orientations and positions of helices in the 7TM bundles are conserved between the two classes (Fig. 1b, Supplementary Fig. 5). The 7TM helices of GCGR superimpose with those of the class A receptors with root mean squared deviation (r.m.s.d.) of C_α backbone atoms in the 2.7–3.3 Å range, above the 2.2–3.0 Å range observed between major branches (α, β, γ and δ) of class A GPCRs. The structural alignment of GCGR with rhodopsin shows an approximate spatial correspondence between residues in the 7TM helices of the two GPCR classes, but also reveals a number of gaps in transmembrane regions reflecting substantial structural deviations in transmembrane helices (Supplementary Fig. 6). The spatial correspondence between 7TM residues makes it possible to project the widely used class A Ballesteros-Weinstein numbering scheme¹⁷ (used hereafter for class A as BW number in parentheses) for comparisons between GPCR classes (Supplementary Table 3). Analysis of sequence and structural features within class B GPCRs, however, is defined by the Wootten numbering scheme based on class B residue conservation¹⁸ (used hereafter for class B receptors as superscript, Supplementary Table 3).

Class B versus A GPCRs

The GCGR structure reveals a number of features in the 7TM domain that are distinct from known class A GPCRs. The N-terminal end of helix I in GCGR is longer than any known class A GPCR structures and extends three additional helical turns (approximately 16 Å) above

¹Department of Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. ²The National Center for Drug Screening and the CAS Key Laboratory of Receptor Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences (CAS), 189 Guo Shou Jing Road, Shanghai, 201203, China. ³Division of Medicinal Chemistry, Faculty of Sciences, Amsterdam Institute for Molecules, Medicines and Systems (AIMMS), VU University of Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands. ⁴The Joint Center for Structural Genomics, Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA. ⁵Protein & Peptide Chemistry, Novo Nordisk, Novo Nordisk Park, 2760 Malov, Denmark.

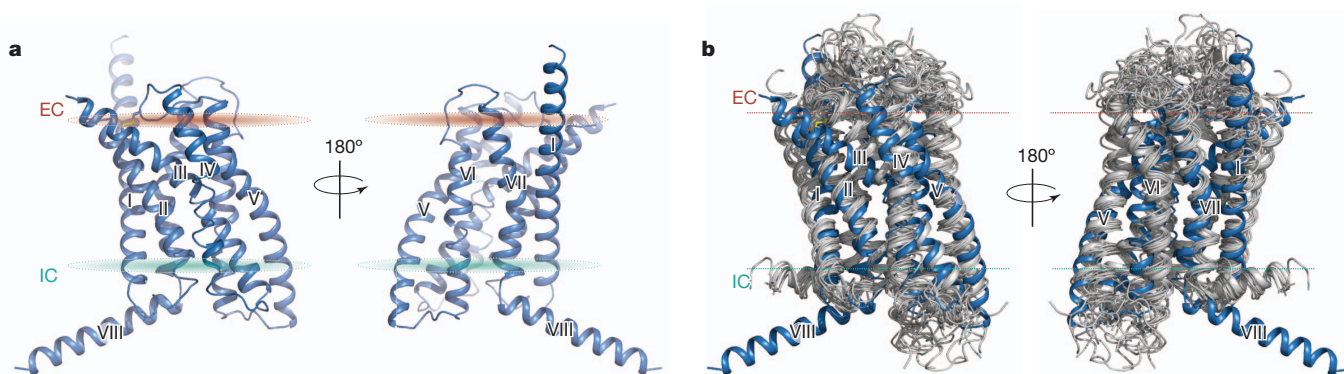


Figure 1 | Structure of the 7TM domain of human GCGR and comparison to class A GPCR structures. **a**, Cartoon depiction of the 7TM domain structure of GCGR. The two views are rotated 180° relative to each other. The disulphide bond between helix III and extracellular loop 2 (ECL2) is shown as yellow sticks. **b**, Side view of structural superimposition of 7TM domains of

the extracellular (EC) membrane boundary from Lys 136 to Gly 125 (Fig. 1). This region of GCGR, referred to as the stalk, may be involved in glucagon binding and helps to define the orientation of the ECD with respect to the 7TM domain. Extracellular loop 1 (ECL1) of GCGR is 16 residues long, as compared with 4–6 residues in most class A GPCRs. Although the ECL1 residues 201–215 are not resolved in the crystal structure, mutagenesis studies presented here and elsewhere^{19–21} indicate that these residues are involved in interactions with peptide ligands. The distance between the EC tips of 7TM helices II and VI is the largest among GPCR structures, and the distance between the EC tips of helices III and VII is among the largest, except for kappa (κ -OR) and mu (μ -OR) opioid receptors^{22,23} (Supplementary Fig. 5). The positioning of the EC tips of these 7TM helices creates a wider and deeper cavity in the ligand-binding pocket of GCGR, which is larger than in any class A receptor structures (Fig. 2, Supplementary Table 4).

At the intracellular (IC) side, the distances between the helical tips of GCGR are within the same range as those in class A structures, except for an extensive inward shift of the IC tip of helix VII (Supplementary Fig. 5). Although the inward shift in the IC part of helix

GCGR (blue) and class A GPCRs (grey). Structures of class A GPCRs used (PDB): 1U19, 2RH1, 2YCW, 3RZE, 3PBL, 3UON, 4DAJ, 3EML, 3V2W, 3ODU, 4DJH, 4EA3, 4DKL, 4EJ4 and 3VW7. Extracellular (EC) and intracellular (IC) membrane boundaries (predicted by OMP server⁴⁴) are shown as brown and cyan ovals (**a**) or dotted lines (**b**), respectively.

VII is a hallmark of class A receptor activation³, it is not yet clear what role the IC region of helix VII plays in GCGR. The receptor lacks a proline kink in helix VII, which is a part of the conserved NP (BW 7.50) xxY motif in class A GPCRs²⁴; instead, helix VII of GCGR has a glycine residue (Gly 393^{7.50}) that allows for a helical bend in this region. This glycine in helix VII of GCGR is part of the FQG^{7.50}xxVxxxY^{7.57}CF motif that is fully conserved in secretin-like class B receptors (Fig. 3a). This Gly 393^{7.50} induced bend is stabilized by hydrophobic interactions with Phe 184^{2.57} of helix II in GCGR (Fig. 3a).

The GCGR structure also includes an IC helix VIII comprising 20 residues at the C-terminal end of the receptor that tilts approximately 25° away from the membrane as compared with its consensus position in class A (Fig. 1). This tilt is probably a result of crystal packing interactions (Supplementary Fig. 4), but it should be noted that Glu 406 in helix VIII is fully conserved in secretin-like class B receptors, and forms two interhelical salt bridges with conserved residues Arg 173^{2.46} and Arg 346^{6.37} (Fig. 3b). Though the tilt of helix VIII may alter interactions in the region, conformational modelling with helix VIII parallel to the membrane suggests that the Glu 406 salt bridges are preserved in this

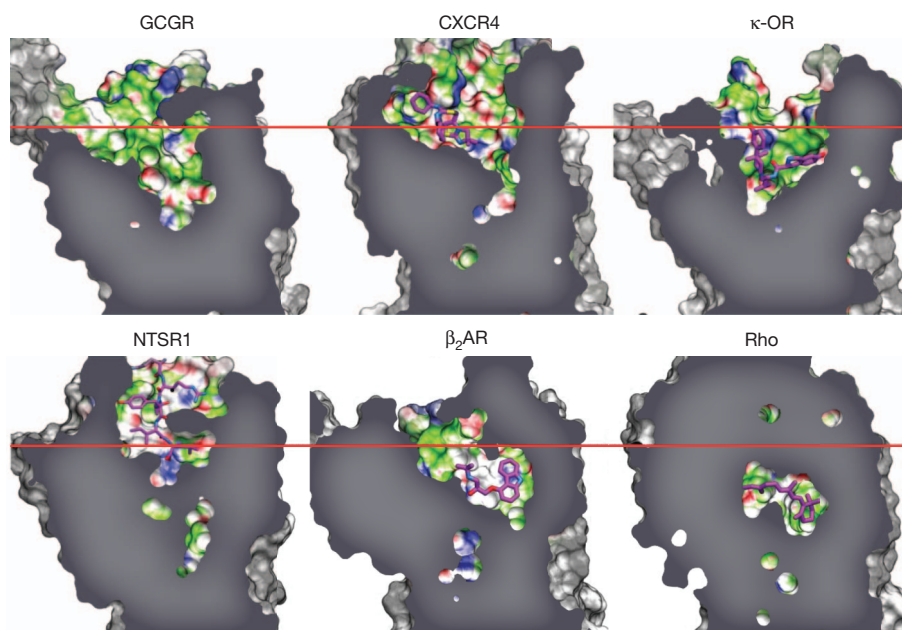


Figure 2 | Comparison of the ligand-binding pocket of GCGR with class A GPCRs. The binding cavity of GCGR is compared with the binding cavities of human chemokine receptor CXCR4 (PDB: 3ODU), human κ -opioid receptor (κ -OR) (PDB: 4DJH), rat neurotensin receptor (NTSR1) (PDB:

4GRV), human β_2 -adrenergic receptor (β_2 AR) (PDB: 2RH1) and bovine rhodopsin (Rho) (PDB: 1U19) (Supplementary Table 4). The approximate position of the EC membrane boundary is shown as a red line and bound ligands as magenta carbon atoms.

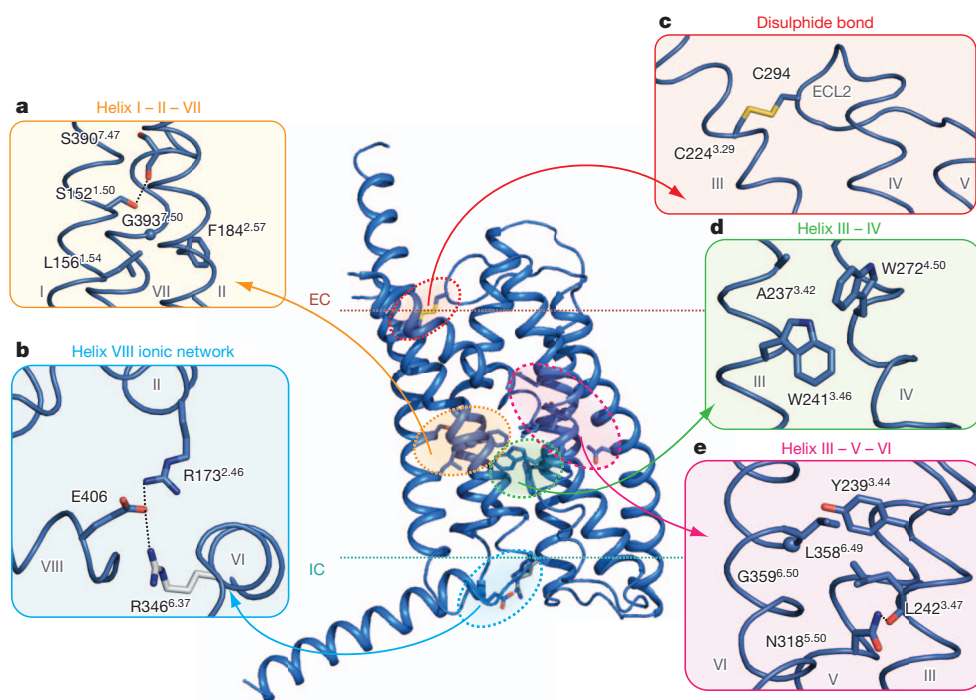


Figure 3 | Structural features of class B GPCRs. Comparison of GCGR and class A GPCR crystal structures indicates distinct and conserved features. **a, d, e,** The homologous GCGR residues involved in helix I–II, III–IV, and III–VI interface interactions as discussed for class A receptors by Venkatakrishnan *et al.*²⁵, and class B GPCR specific residues that mediate helix I–VII, II–VII, and III–V interface interactions. **b,** GCGR residues Glu 406 of helix VIII, Arg 173^{2,46}, and Arg 346^{6,37} form a class B receptor specific ionic network. Arg 346^{6,37} (grey) has a weak electron density. **c,** The disulphide bond between Cys 224^{3,29} and Cys 294 of ECL2 in the GCGR structure is a conserved feature between classes A and B receptors. Hydrogen bond interactions and salt bridges are indicated by black dashed lines. Electron density maps for residues in this figure are shown in Supplementary Fig. 10. Comparison of class A Ballesteros–Weinstein and class B Wootten residue numbering is provided in Supplementary Table 3.

conformation and are likely to be a distinct feature of secretin-like class B receptors because there is no strong conservation among these residues in class A.

The GCGR 7TM structure also reveals several structural features that are conserved between class A and B receptors. One such feature is a disulphide bond between Cys 294 in ECL2 and Cys 224^{3,29} (BW 3.25, Supplementary Fig. 6), which apparently stabilizes the receptor's 7TM fold (Fig. 3c). Another conserved feature of a common GPCR fold²⁵ involves similar regions of contacts between helices I–II, I–VII, III–IV and III–VI in class A and B GPCRs. The two GPCR classes, however, contain different patterns of conserved residues in these positions (Fig. 3, Supplementary Fig. 6). In class B GPCRs, the helix I–II interaction is stabilized by conserved hydrophobic residues Leu 156^{1,54} and Phe 184^{2,57}, class A GPCRs contain conserved polar residues Asn (BW 1.50) and Asp (BW 2.50) in this region²⁵. At the helix I–VII interface, Ser 152^{1,50} forms a hydrogen bond with the backbone of Ser 390^{7,47} (Fig. 3a). Mutation of homologous glucagon-like peptide-1 receptor (GLP1R) residues Ser 155^{1,50} and Ser 392^{7,47} alters receptor signalling¹⁸. At the GCGR helix III–IV interface (Fig. 3d), the conserved residue Trp 272^{4,50} interacts with Trp 241^{3,46}, whereas in class A structures the Trp residue in helix IV (BW 4.50) interacts with the residue at BW position 3.38 in helix III (homologous GCGR residue Ala 237^{3,42}) (refs 24, 25). The helix III–VI interface (Fig. 3e) in secretin-like class B GPCRs contains conserved hydrophobic residues Tyr 239^{3,44} (or Phe^{3,44}) and Leu 358^{6,49} (or Phe^{6,49}) which make similar hydrophobic interactions as structurally aligned Ile/Val/Leu (BW 3.40) and Phe (BW 6.44) residues present in most class A GPCRs²⁵ (Supplementary Fig. 6). This interface is further stabilized in class B GPCRs by close contact between the conserved Tyr 239^{3,44} and Gly 359^{6,50}. Another class B GPCR specific interhelical hydrogen bond is formed between the conserved Asn 318^{5,50} and the backbone of Leu 242^{3,47} at the helix III–V interface (Fig. 3e).

Recognition between GCGR and glucagon

To better understand GCGR–glucagon interactions, we performed a comprehensive mutagenesis and glucagon-binding study of GCGR at 90 different residue positions (Fig. 4, Supplementary Table 5). A total

of 129 mutants were tested, and of these, 110 covering 85 different positions had expression levels greater than 30% of wild-type GCGR. Of them, 41 mutations covering 28 different positions in the GCGR 7TM domain had more than fourfold reduction in glucagon binding (IC₅₀ values) relative to wild-type GCGR. The results of these GCGR mutation studies were mapped onto the crystal structure of the GCGR 7TM domain (Fig. 5). Most of the residues that play an important role in glucagon binding face the main cavity in the 7TM core, and form a binding site that covers parts of ECL1, ECL2 and ECL3 and helices I, II, III, V, VI and VII, and extends deep into the 7TM cavity.

To investigate the recognition between glucagon and GCGR, we built a glucagon-bound GCGR structure model, based on the GCGR 7TM domain crystal structure, the GCGR ECD structure (Protein Data Bank (PDB) accession: 4ERS)⁶, the ECD structure of the GCGR homologue GLP1R bound to the GLP1 peptide (PDB: 3IOL)⁸, and the N-capped conformation of pituitary adenylate cyclase activating polypeptide (PACAP; PDB: 1GEA)²⁶ (Fig. 5a). The model further included several experimentally supported distance restraints between GCGR and glucagon based on photo-crosslinking studies between GLP1R and GLP1 (ref. 12).

The predicted binding mode of glucagon to the ECD of GCGR (Fig. 5b) is in line with our results (Fig. 4, Supplementary Table 5) and previously reported mutation studies on GCGR^{6,27} and GLP1R⁸. Figure 5b shows how GCGR residues Asp 63, Tyr 65 and Lys 98 function in stabilizing the ECD as observed in the GCGR ECD crystal structure⁶ and supported by mutagenesis studies^{6,27} (Fig. 4, Supplementary Table 5). The Trp 36 side chain is an important hydrophobic interaction site for the C-terminal region of glucagon, similar to Trp 39 in the GLP1–GLP1R-ECD crystal structure⁸. The stalk, observed in helix I of the GCGR 7TM crystal structure, links the ECD and 7TM domain in the model (Figs 5a, b). The α -helical conformation of the stalk is supported by intrahelical interactions in the crystal structure (Glu 133–Lys 136) and model (Glu 127–Gln 131 and Glu 129–Lys 132), and is likely to be further stabilized by interactions with the extended ECL1 and the α -helical portion of glucagon. The potential function of the α -helical stalk in glucagon binding is supported by a reduced glucagon affinity to the Ala 135 Pro mutant (Fig. 4d, Supplementary Table 5), which probably distorts the α -helical

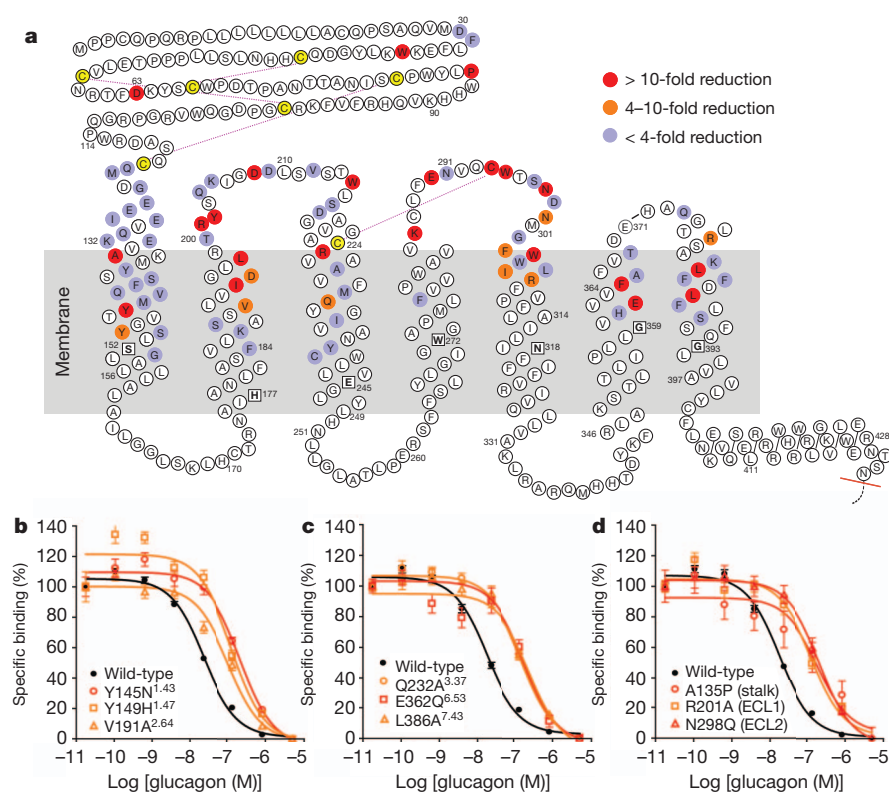


Figure 4 | Effects of mutation studies in GCGR snake plot. **a**, Mutated residues that show <4-fold (purple), 4–10-fold (orange), and >10-fold (red) changes of IC_{50} values for glucagon binding with receptor expression >30% of wild-type (Supplementary Table 5). Mutation studies to investigate peptide ligand binding have been previously reported for several class B GPCRs including GCGR^{6,19,21,27,31,32,34}, GLP1R^{6,8,18,20,29,36}, GIPR^{3,37}, rSCTR^{11,30,35,38,43} and VPAC1 (refs 39–41) (Supplementary Table 6). The most conserved residues in helices I to VII of class B GPCRs¹⁸ are boxed and shown in bold. **b–d**, Representative binding curves of GCGR mutants with glucagon. Data are expressed as a percentage of specific ^{125}I -glucagon binding in the presence of 0.02 nM unlabelled peptide. Each point (\pm s.e.m.) represents the mean value of at least three independent experiments done in triplicate (IC_{50} are shown in Supplementary Table 5).

conformation of the stalk region. The stalk may also function as part of the binding site for the previously proposed middle hinge region of glucagon²⁸. An α -helical conformation in this region of GLP1R in complex with the GLP1 peptide was recently proposed based on cross-linking data between receptor and peptide residues¹². In the GCGR–glucagon model, the corresponding pairs of residues, F6–Gln 142^{1.40} and Y10–Tyr 138^{1.36} (one letter amino acid abbreviation is used to designate glucagon residues), are located in close proximity and point towards each other, supporting a similar interaction mode as proposed in the GLP1R–GLP1 complex¹² (Fig. 5b). The 12 Å distance between L14–Trp 295^{ECL2} in the GCGR–glucagon model exceeds the range of crosslinking distances in previous GLP1R–GLP1 models (8–9 Å)¹², though this may reflect differences between GCGR and GLP1R ligand-binding modes.

There is no clear consensus on the binding site location of peptide ligands in the 7TM domain of class B GPCRs, which has been associated either with the ECL regions⁴⁷, or with a pocket in the 7TM domain^{11,12,29}. The GCGR–glucagon model illustrates a way to account for the extensive interactions of the peptide with ECLs, as well as residues deep in the 7TM domain (Fig. 5c, d). First, the GCGR crystal structure reveals that some of the binding site residues previously positioned at the top of 7TM helices or in ECLs^{30–32} are in fact located deeper in the 7TM domain. Second, an extended flexible conformation of the first five residues allows glucagon to reach deep into the pocket. Our model of glucagon incorporates a hypothetical N-capping conformation³³ of the peptide helix in residues F6–T7–Y10, similar to the one observed in the receptor-bound PACAP²⁶, though other conformations of this region are possible.

Most interactions predicted by the GCGR–glucagon binding model are supported by mutagenesis (Figs 4, 5c, d, Supplementary Table 5) and photo-crosslinking studies on GCGR^{6,19,21,27,31,32,34} and other class B GPCRs^{8–13,20,29,30,35,36}. In the 7TM domain, many residues predicted to interact with glucagon show dramatic effects on glucagon binding without reducing receptor expression (Fig. 4, Supplementary Table 5). Figure 5 shows how these mutations line the 7TM binding site in the GCGR–glucagon model and include residues that are located deep

in the pocket (Tyr 149^{1.47}, Val 191^{2.64}, Gln 232^{3.37}, Glu 362^{6.53} and Leu 386^{7.43}). These results strongly support extension of the N terminus of glucagon deep into the GCGR pocket, a region that could be equally important for ligand binding as in class A GPCRs.

In the loop region, residues Arg 201, Tyr 202, Asp 208 and Trp 215 of GCGR either stabilize the ECL1 conformation and/or directly interact with glucagon (Fig. 5c, d). The GCGR–glucagon binding model further suggests that residues Trp 295 and Asn 298 directly interact with glucagon, as mutation of these ECL2 residues strongly affects ligand binding. Although mutations of Asp 218 (ref. 21), Cys 224^{3,29} (refs 21, 27), Arg 225^{3,30}, Lys 286^{4,64}, Glu 290 and Cys 294 (Fig. 4, Supplementary Table 5) also affect ligand binding, these residues do not directly interact with glucagon in the model, but can play a role in stabilizing the loop conformation compatible with glucagon binding. For example, ECL2 and ECL1 are stabilized by a disulphide bridge between Cys 294 and Cys 224^{3,29}, and potential salt bridges between Lys 286^{4,64} and Glu 290, and between Asp 218 and Arg 225^{3,30}, respectively. Similarly, Arg 378 is proposed to play a role in glucagon binding indirectly by stabilizing the ECL3 conformation, while Trp 304^{5,36} stabilizes ECL2 at the interface between helices V and VI.

The GCGR–glucagon model presented in Fig. 5 is based on crystallographic evidence and is consistent with the results of extensive mutation binding studies (Fig. 4, Supplementary Table 5)^{6,19,21,27,31,32,34}, and thus provides comprehensive insight into recognition between the native ligand and its receptor. The hypothetical model of the complex can offer a useful platform for the design of biochemical and biophysical experiments detailing the complex structure, as well as the design of stabilized constructs that may lead to solution of the full-length receptor–ligand complex.

Related class B GPCRs

The GCGR–glucagon model can be informative for understanding common features that determine ligand recognition of other class B receptors. The GCGR mutation data (Fig. 4, Supplementary Table 5), and previous studies^{6,19,21,27,31,32,34} are paralleled by mutagenesis of homologous residues in other class B GPCRs. Supplementary Table 6 shows

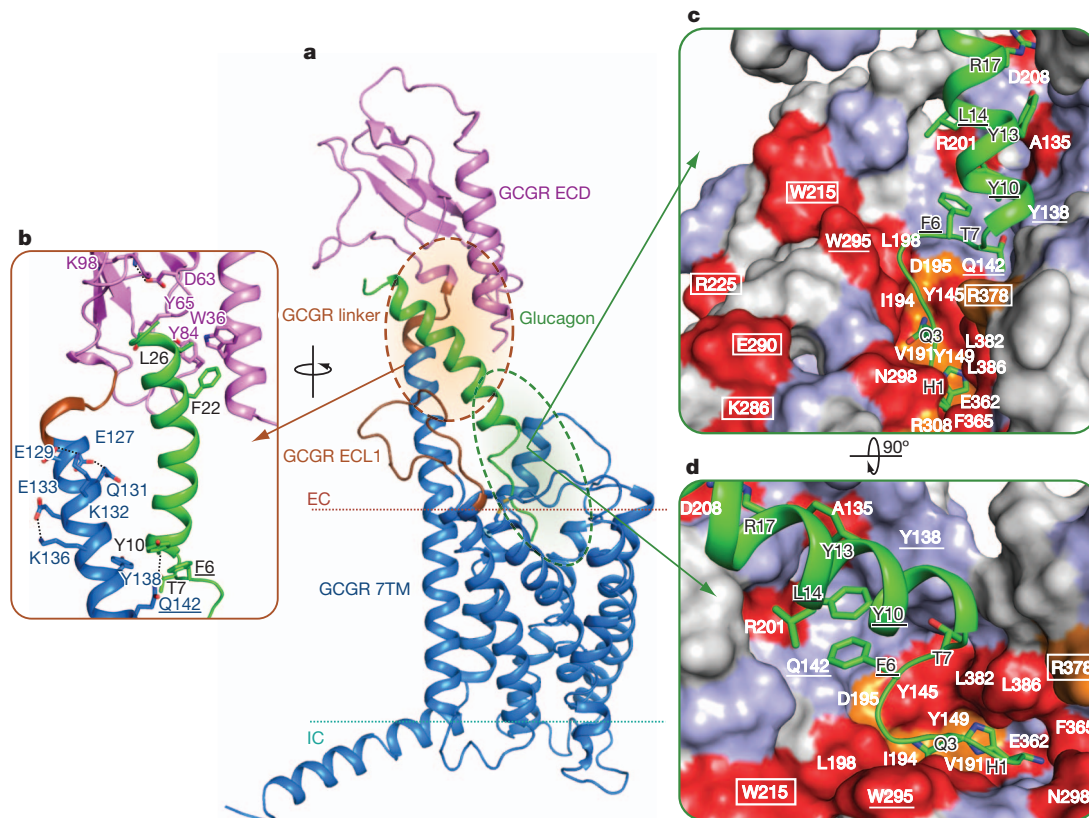


Figure 5 | Model of GCGR bound to glucagon. **a, b,** GCGR with the ECD (magenta) and 7TM domain (blue) bound to glucagon (green). Residues 122–126 and 199–218 (brown) are not defined in the GCGR ECD (GCGR-linker) (PDB: 4ERS) and 7TM domain (ECL1) crystal structures, respectively. The GCGR ECD structure and the interactions between GCGR ECD and glucagon resemble those in the GCGR ECD (PDB: 4ERS)⁶ and GLP1R–GLP1R-ECD complex (PDB: 3IOL)⁸ structures, respectively. **c, d,** The effects of mutation

an overview of 274 previously reported mutants of GCGR^{6,19,21,27,31,32,34}, GLP1R^{8,18,20,29,36}, glucose-dependent insulinotropic polypeptide receptor (also known as gastric inhibitory polypeptide receptor) (GIPR)^{9,37}, rat secretin receptor (rSCTR)^{11,30,35,38,43}, and vasoactive intestinal peptide and pituitary adenylate cyclase-activating polypeptide receptor 1 (VPAC1, also known as VIPR1) (refs 39–41). For example, mutations of other class B GPCRs in residues that align to GCGR residues Tyr 65^{ECD} (ref. 8), Tyr 84^{ECD} (ref. 8), Leu 85^{ECD} (ref. 8), Tyr 145^{1.43} (ref. 35), Tyr 149^{1.47} (refs 29, 35), Lys 187^{2.60} (refs 9, 18, 20, 29, 30, 39), Ile 194^{2.67} (refs 9, 18, 20, 29, 30, 39), Asp 195^{2.68} (refs 11, 20, 29, 30), Leu 198^{2.71} (refs 11, 21), Arg 225^{3.30} (ref. 20), Gln 232^{3.37} (refs 9, 29), Lys 286^{4.64} (ref. 36), Glu 290^{ECL2} (ref. 36), Trp 295^{ECL2} (refs 11, 36), Asn 298^{ECL2} (refs 11, 36), Phe 365^{5.56} (refs 9, 11) and Leu 386^{7.43} (ref. 40) have been shown to affect peptide ligand binding and/or potency, supporting the GCGR–glucagon model in Fig. 5. The GCGR–glucagon model demonstrates that residues which have been identified to interact with the homologous residues Q3 of glucagon, D3 of secretin, and D3 of vasoactive intestinal peptide are located within the same vicinity in the 7TM domain of GCGR (Lys 187^{2.60} and Ile 194^{2.67}) (refs 31, 32), rSCTR (Tyr 128^{1.47}, Arg 166^{2.60}, Lys 173^{2.67}, and Asp 174^{2.68}) (refs 30, 35), and VPAC1 (Arg 188^{2.60} and Lys 195^{2.67}) (ref. 39), respectively (Supplementary Table 6).

The distinct structural features and larger binding pocket of the GCGR 7TM domain provide new insights into the molecular details of peptide ligand binding, and a more reliable structural template for the design of specific and potent small molecules for the treatment of type 2 diabetes. Moreover, the apparent overlap of class B GPCR binding sites suggests that, despite possible structural differences

studies of individual GCGR residues on glucagon (green) binding mapped onto the GCGR binding surface using the colour coding presented in Fig. 4. Important glucagon residues are labelled black. GCGR residues proposed to be important in stabilizing extracellular loops are boxed. GCGR–glucagon residue pairs that are homologous to residue pairs identified in GLP1R–GLP1 crosslinking studies¹² are underlined.

between class B GPCRs, the GCGR crystal structure might offer new opportunities to construct structural models to describe interactions between peptide ligands and other class B GPCRs. This is particularly exciting for those receptors involved in glucose regulation, including GLP1R and GIPR.

METHODS SUMMARY

BRIL–GCGR(ΔECD/ΔC) was expressed in *Spodoptera frugiperda* insect cells, solubilized with 1/0.1% (w/v) of *n*-dodecyl-β-D-maltopyranoside and cholesteryl hemisuccinate for 2 h at 4 °C, and purified by immobilized metal ion affinity chromatography with 50–200 μM of GCGR antagonist ligand NNC0640. Protein at 80 mg ml^{−1} was mixed with monoolein and cholesterol in a ratio of 40%:54%:6% (w/w/w) to form lipidic cubic phase⁴², and crystallized in 100 mM MES pH 6.0, 140–200 mM Na/K tartrate tetrahydrate, 9–17% (v/v) PEG 400, 0.35–0.55% (v/v) Jeffamine M-600, pH 7.0, and 200 μM NNC0640 at 20 °C. X-ray data were collected on the 23ID-D beamline (GM/CA CAT) at the Advanced Photon Source (Argonne, Illinois) using a 20-μm minibeam at wavelength of 1.0330 Å (Supplementary Table 2). A single wavelength anomalous dispersion (SAD) data set was collected at 4 Å from a single crystal soaked with tantalum bromide (Ta₆Br₁₂; Supplementary Fig. 7). Phase information from the SAD data set confirmed the molecular replacement solution obtained from an auto script that used mixed models of all known class A GPCR structures as search models (Supplementary Fig. 8). Native diffraction data were collected from 14 crystals and anisotropically truncated before refinement in a*, b*, and c* to 3.3, 3.4, and 3.3 Å, respectively. We report the final structure at 3.4 Å resolution, and data collection, processing, structure solution and refinement are described in the Methods.

The model of the GCGR–glucagon complex was constructed using the structures of the GCGR 7TM domain presented here, the GCGR ECD (PDB: 4ERS), the GLP1R–GLP1 complex (PDB: 3IOL), and the N-capped conformation of PACAP (PDB: 1GEA).

Binding studies were performed using transiently transfected CHO-K1 and HEK293T cells. Either whole cells or prepared membranes were used to measure binding affinity (IC_{50}) of glucagon or NNC0640 using radiolabelled glucagon or NNC0640.

Full Methods and any associated references are available in the online version of the paper.

Received 7 March; accepted 17 June 2013.

Published online 17 July 2013.

- Lagerström, M. C. & Schiöth, H. B. Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nature Rev. Drug Discov.* **7**, 339–357 (2008).
- Cho, Y. M., Merchant, C. E. & Kieffer, T. J. Targeting the glucagon receptor family for diabetes and obesity therapy. *Pharmacol. Ther.* **135**, 247–278 (2012).
- Katritch, V., Cherezov, V. & Stevens, R. C. Structure-function of the G protein-coupled receptor superfamily. *Annu. Rev. Pharmacol. Toxicol.* **53**, 531–556 (2013).
- Hoare, S. R. Mechanisms of peptide and nonpeptide ligand binding to class B G-protein-coupled receptors. *Drug Discov. Today* **10**, 417–427 (2005).
- Pal, K., Melcher, K. & Xu, H. E. Structure and mechanism for recognition of peptide hormones by Class B G-protein-coupled receptors. *Acta Pharmacol. Sin.* **33**, 300–311 (2012).
- Koth, C. M. et al. Molecular basis for negative regulation of the glucagon receptor. *Proc. Natl Acad. Sci. USA* **109**, 14393–14398 (2012).
- Parthier, C., Reedtz-Runge, S., Rudolph, R. & Stubbs, M. T. Passing the baton in class B GPCRs: peptide hormone activation via helix induction? *Trends Biochem. Sci.* **34**, 303–310 (2009).
- Underwood, C. R. et al. Crystal structure of glucagon-like peptide-1 in complex with the extracellular domain of the glucagon-like peptide-1 receptor. *J. Biol. Chem.* **285**, 723–730 (2010).
- Yaqub, T. et al. Identification of determinants of glucose-dependent insulinotropic polypeptide receptor that interact with N-terminal biologically active region of the natural ligand. *Mol. Pharmacol.* **77**, 547–558 (2010).
- Miller, L. J., Dong, M., Harikumar, K. G. & Gao, F. Structural basis of natural ligand binding and activation of the Class II G-protein-coupled secretin receptor. *Biochem. Soc. Trans.* **35**, 709–712 (2007).
- Dong, M. et al. Mapping spatial approximations between the amino terminus of secretin and each of the extracellular loops of its receptor using cysteine trapping. *FASEB J.* **26**, 5092–5105 (2012).
- Miller, L. J. et al. Refinement of glucagon-like peptide 1 docking to its intact receptor using mid-region photolabile probes and molecular modeling. *J. Biol. Chem.* **286**, 15895–15907 (2011).
- Dong, M. et al. Molecular basis of secretin docking to its intact receptor using multiple photolabile probes distributed throughout the pharmacophore. *J. Biol. Chem.* **286**, 23888–23899 (2011).
- Gensure, R. C., Shimizu, N., Tsang, J. & Gardella, T. J. Identification of a contact site for residue 19 of parathyroid hormone (PTH) and PTH-related protein analogs in transmembrane domain two of the type 1 PTH receptor. *Mol. Endocrinol.* **17**, 2647–2658 (2003).
- de Graaf, C., Rein, C., Piwnicka, D., Giordanetto, F. & Rognan, D. Structure-based discovery of allosteric modulators of two related class B G-protein-coupled receptors. *ChemMedChem* **6**, 2159–2169 (2011).
- Chun, E. et al. Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure* **20**, 967–976 (2012).
- Ballesteros, J. A. & Weinstein, H. In *Methods in Neurosciences* Vol. 25 (ed. Sealfon, S. C.) 366–428 (Academic Press, 1995).
- Wooten, D., Simms, J., Miller, L. J., Christopoulos, A. & Sexton, P. M. Polar transmembrane interactions drive formation of ligand-specific and signal pathway-biased family B G protein-coupled receptor conformations. *Proc. Natl Acad. Sci. USA* **110**, 5211–5216 (2013).
- Unson, C. G. et al. Roles of specific extracellular domains of the glucagon receptor in ligand binding and signaling. *Biochemistry* **41**, 11795–11803 (2002).
- Xiao, Q., Jeng, W. & Wheeler, M. B. Characterization of glucagon-like peptide-1 receptor-binding determinants. *J. Mol. Endocrinol.* **25**, 321–335 (2000).
- Roberts, D. J., Vertongen, P. & Waelbroeck, M. Analysis of the glucagon receptor first extracellular loop by the substituted cysteine accessibility method. *Peptides* **32**, 1593–1599 (2011).
- Wu, H. et al. Structure of the human κ -opioid receptor in complex with JDTic. *Nature* **485**, 327–332 (2012).
- Manglik, A. et al. Crystal structure of the μ -opioid receptor bound to a morphinan antagonist. *Nature* **485**, 321–326 (2012).
- Fredriksson, R., Lagerström, M. C., Lundin, L. G. & Schiöth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256–1272 (2003).
- Venkatakrishnan, A. J. et al. Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185–194 (2013).
- Inooka, H. et al. Conformation of a peptide ligand bound to its G-protein coupled receptor. *Nature Struct. Biol.* **8**, 161–165 (2001).
- Prévost, M. et al. Mutational and cysteine scanning analysis of the glucagon receptor N-terminal domain. *J. Biol. Chem.* **285**, 30951–30958 (2010).
- Ahn, J. M., Medeiros, M., Trivedi, D. & Hruby, V. J. Development of potent truncated glucagon antagonists. *J. Med. Chem.* **44**, 1372–1379 (2001).
- Coopman, K. et al. Residues within the transmembrane domain of the glucagon-like peptide-1 receptor involved in ligand binding and receptor activation: modelling the ligand-bound receptor. *Mol. Endocrinol.* **25**, 1804–1818 (2011).
- Di Paolo, E. et al. Contribution of the second transmembrane helix of the secretin receptor to the positioning of secretin. *FEBS Lett.* **424**, 207–210 (1998).
- Perret, J. et al. Mutational analysis of the glucagon receptor: similarities with the vasoactive intestinal peptide (VIP)/pituitary adenylate cyclase-activating peptide (PACAP)/secretin receptors for recognition of the ligand's third residue. *Biochem. J.* **362**, 389–394 (2002).
- Runge, S. et al. Three distinct epitopes on the extracellular face of the glucagon receptor determine specificity for the glucagon amino terminus. *J. Biol. Chem.* **278**, 28005–28010 (2003).
- Neumann, J. M. et al. Class-B GPCR activation: is ligand helix-capping the key? *Trends Biochem. Sci.* **33**, 314–319 (2008).
- Cascieri, M. A. et al. Characterization of a novel, non-peptidyl antagonist of the human glucagon receptor. *J. Biol. Chem.* **274**, 8694–8697 (1999).
- Di Paolo, E. et al. Mutations of aromatic residues in the first transmembrane helix impair signalling by the secretin receptor. *Receptors Channels* **6**, 309–315 (1999).
- Koole, C. et al. Second extracellular loop of human glucagon-like peptide-1 receptor (GLP-1R) has a critical role in GLP-1 peptide binding and receptor activation. *J. Biol. Chem.* **287**, 3642–3658 (2012).
- Tseng, C. C. & Lin, L. A point mutation in the glucose-dependent insulinotropic peptide receptor confers constitutive activity. *Biochem. Biophys. Res. Commun.* **232**, 96–100 (1997).
- Ganguli, S. C. et al. Protean effects of a natural peptide agonist of the G protein-coupled secretin receptor demonstrated by receptor mutagenesis. *J. Pharmacol. Exp. Ther.* **286**, 593–598 (1998).
- Solano, R. M. et al. Two basic residues of the h-VPAC1 receptor second transmembrane helix are essential for ligand binding and signal transduction. *J. Biol. Chem.* **276**, 1084–1088 (2001).
- Ceraudo, E. et al. Spatial proximity between the VPAC1 receptor and the amino terminus of agonist and antagonist peptides reveals distinct sites of interaction. *FASEB J.* **26**, 2060–2071 (2012).
- Tan, Y. V., Couvineau, A. & Laburthe, M. Diffuse pharmacophoric domains of vasoactive intestinal peptide (VIP) and further insights into the interaction of VIP with the N-terminal ectodomain of human VPAC1 receptor by photoaffinity labeling with [Bpa6]-VIP. *J. Biol. Chem.* **279**, 38889–38894 (2004).
- Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols* **4**, 706–731 (2009).
- Di Paolo, E. et al. Role of charged amino acids conserved in the vasoactive intestinal polypeptide/secretin family of receptors on the secretin receptor functionality. *Peptides* **20**, 1187–1193 (1999).
- Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by NIH Roadmap grant P50GM073197 for technology development (V.C. and R.C.S.), and PSI:Biological grant U54 GM094618 for biological studies and structure production (target GPCR-49) (V.K., V.C. and R.C.S.); PSI:Biological grant U54 GM094586 for structure QC; The Ministry of Health grants 2012ZX09304-011 and 2013ZX09507002 (M.-W.W.), Shanghai Science and Technology Development Fund 11DZ2292200 (M.-W.W.); Novo Nordisk-Chinese Academy of Sciences Research Fund NNCAS-2011-7 (M.-W.W.); Thousand Talents Program in China (R.C.S. and M.-W.W.); NIH Postdoctoral Training Grant (NRS) F32 DK088392 (F.Y.S.); The Netherlands Organization for Scientific Research (NWO) through a VENI grant (Grant 700.59.408 to C.d.G.); COST Action CM1207, GLISTEN (C.d.G.). We also thank V. Hruby and M. Cai for advice with the glucagon binding assay and general discussions; J. Velasquez for help with molecular biology; T. Trinh and M. Chu for help with baculovirus expression; K. Kadyshkevskaya for assistance with figure preparation; X. Q. Cai, J. Wang, Y. Feng, A. T. Dai, Y. Zhou, J. J. Deng, Y. B. Dai and J. W. Zhao for technical assistance in mutation studies; A. Walker for assistance with manuscript preparation; and J. Smith and R. Fischetti for assistance in development and use of the minibeam and beamtime at GM/CA-CAT beamline 23-ID at the Advanced Photon Source, which is supported by National Cancer Institute grant Y1-CO-1020 and National Institute of General Medical Sciences grant Y1-GM-1104.

Author Contributions F.Y.S. designed, expressed, characterized and screened constructs and ligands for crystallization. F.Y.S. purified and crystallized the receptor in LCP, optimized crystallization conditions, grew crystals, collected diffraction data and prepared the manuscript. G.W.H. and Q.X. solved and refined the structure, and prepared the manuscript. V.C. collected and processed diffraction data, and prepared the manuscript. M.H., D.Y., Z.Z. and C.Z. expressed the receptor, and performed the mutagenesis and ligand-binding assay. V.K. and C.d.G. designed and analysed the receptor mutagenesis studies, constructed the receptor–ligand model and prepared the manuscript. D.W. and J.S.J. collected and processed SAD data and determined an initial electron density map from experimental phases. W.L. and V.C. trained and assisted in LCP crystallization. J.L. provided ligands for GCGR and prepared the manuscript. R.C.S., F.Y.S., M.-W.W., V.K., V.C. and C.d.G. were responsible for the overall project strategy and management and wrote the manuscript.

Author Information The coordinates and the structure factors have been deposited in the Protein Data Bank under the accession code 4L6R. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.C.S. (stevens@scripps.edu) or M.-W.W. (wangmw@mail.shncn.ac.cn).

METHODS

BRIL–GCGR(ΔECD/ΔC) construct design and S9 expression. The human wild-type GCGR DNA was synthesized by DNA 2.0 and codon optimized for expression in *Spodoptera frugiperda* (S9) insect cells. The BRIL–GCGR(ΔECD/ΔC) fusion construct was made by deleting N-terminal residues 1–122, fusing the thermally stabilized apocytocrome b₅₆₂RIL (M7W, H102I, R106L) (referred to as BRIL) from *E. coli* at residue 123 (ref. 16), and truncating the C terminus at residue 432 to create the final construct for crystallization (Supplementary Fig. 2). This chimaeric construct was obtained after screening 60 constructs of different BRIL junction and C-terminal truncation sites to generate crystals with diffraction data of the highest quality and resolution. The construct was cloned into a modified pFastBac1 vector (Invitrogen) containing an expression cassette with haemagglutinin signal sequence at the N terminus, and a PreScission protease site, 10×His, and Flag tag at the C terminus. The BRIL–GCGR(ΔECD/ΔC) fusion construct was expressed in S9 cells using the Bac-to-Bac baculovirus expression system as described previously⁴⁵. S9 cells at a density of 2×10^6 – 3×10^6 cells ml^{−1} were infected with P1 or P2 virus at a multiplicity of infection (m.o.i.) of 7.5. Cells were harvested less than 48 h post-infection and cell pellets were stored at −80 °C until used.

BRIL–GCGR(ΔECD/ΔC) fusion construct purification. S9 membranes were prepared with 1 wash cycle of hypotonic buffer (25 mM HEPES, pH 7.5, 10 mM MgCl₂, 20 mM KCl) in the presence of EDTA-free protease inhibitor cocktail tablets (Roche) and 4 wash cycles of high-salt buffer (25 mM HEPES, pH 7.5, 1 M NaCl, 10 mM MgCl₂, 20 mM KCl). Membrane pellets were homogenized in storage buffer (25 mM HEPES, pH 7.5, 500 mM NaCl, 40% glycerol), flash frozen in liquid nitrogen and stored at −80 °C until use.

The GCGR antagonist ligand NNC0640 (Supplementary Fig. 1b) was essential for purification and crystallization of the BRIL–GCGR(ΔECD/ΔC) fusion construct. Two grams of washed membranes containing the BRIL–GCGR(ΔECD/ΔC) fusion construct were resuspended in 30 ml of buffer (25 mM HEPES, pH 7.0, 166 mM NaCl, 13.3% glycerol) and incubated with 270 μM of compound NNC0640 for 30 min at room temperature. The receptor was solubilized with 1/0.1% (w/v) of *n*-dodecyl-β-D-maltopyranoside (Anatrace) and cholesteryl hemisuccinate (Sigma) (DDM/CHS) for 2 h at 4 °C. The insoluble material was pelleted by ultracentrifugation in a Ti70 rotor at 504,300g for 30 min at 4 °C. The NaCl and DDM/CHS concentrations of the supernatant were adjusted to 800 mM and 0.5/0.05%, respectively, by adding equal volume of talon binding buffer (25 mM HEPES, pH 7.0, 1.475 M NaCl, 10% glycerol). Protein was bound to 2 ml of talon superflow resin slurry (Clontech) overnight at 4 °C on a rotator in the presence of 15 mM imidazole, pH 7.5, and 100 μM NNC0640.

The talon resin was washed with 10 × bed volume of wash buffer 1 (25 mM HEPES, pH 7.0, 800 mM NaCl, 10% glycerol, 0.04/0.008% DDM/CHS, 30 μM NNC0640, 40 mM imidazole, pH 7.5). Detergent concentration was lowered by washing the resin with 20 × bed volume of wash buffer 2 (25 mM HEPES, pH 7.0, 500 mM NaCl, 10% glycerol, 0.02/0.004% DDM/CHS, 30 μM NNC0640). The BRIL–GCGR(ΔECD/ΔC) fusion construct was eluted with 2.5 ml of elution buffer (25 mM HEPES, pH 7.0, 150 mM NaCl, 10% glycerol, 0.02/0.004% DDM/CHS, 30 μM NNC0640, 300 mM imidazole, pH 7.5). The eluted BRIL–GCGR(ΔECD/ΔC) fusion construct was desalted with a PD-10 desalting column (GE Healthcare) to remove the imidazole. His-tagged PreScission protease was added to the samples and incubated overnight at 4 °C to remove the C-terminal 10×His and Flag tags. Reverse talon purification was performed to isolate the cleaved BRIL–GCGR(ΔECD/ΔC) fusion construct by flowing the sample through 200 μl talon superflow resin twice. The flow-through material, containing the cleaved BRIL–GCGR(ΔECD/ΔC) fusion construct, was concentrated to 80 mg ml^{−1} using a Vivaspin centrifuge concentrator (GE Healthcare) with a 100 kilodalton (kDa) molecular weight cut-off.

BRIL–GCGR(ΔECD/ΔC) fusion construct lipidic cubic phase (LCP) crystallization. For LCP crystallization, the BRIL–GCGR(ΔECD/ΔC) construct was mixed with molten lipid at a ratio of 40/60% (v/v) using a mechanical syringe mixer⁴². Due to the high detergent concentration, 10–15% (volume of LCP) of 5 M NaCl was added after the lipid and protein were mixed to convert a destabilized lipidic mesophase into LCP⁴⁶. The host lipid for the LCP reconstitution was monoolein (Sigma) with 10% (w/w) of cholesterol (AvantiPolar Lipids). Crystallization trials were set up as previously described⁴⁷. LCP-FRAP was used to identify initial crystallization conditions that led to GCGR crystals⁴⁸. Crystals were obtained at 20 °C in 100 mM MES, pH 6.0, 140–200 mM Na/K tartrate tetrahydrate, 9–17% (v/v) PEG 400, 0.35–0.55% (v/v) Jeffamine M-600, pH 7.0, 200 μM NNC0640; grown to a final size of 50–100 μm in the longest dimension in about 5 days (Supplementary Fig. 3); and harvested from the LCP matrix using 50 μm MiTeGen micromounts and immediately flash frozen in liquid nitrogen. **Data collection and processing.** X-ray data were collected at the 23ID-D beamline (GM/CA CAT) at the Advanced Photon Source (Argonne, Illinois) using a 20-μm minibeam at a wavelength of 1.0330 Å and a MarMosaic 300 charge-coupled device

(CCD) detector. Crystals were aligned and data collected using a strategy similar to other GPCR structures⁴⁹. Typically, 10–15 frames at 1° oscillation and 1–2 s exposure with non-attenuated beam were collected per crystal due to the fast onset of radiation damage. A 93.9% complete at 3.3 Å data set was obtained by indexing, integrating, scaling and merging data sets from 14 crystals using HKL2000 (ref. 50). Analysis of the final data set by the UCLA diffraction anisotropy server (<http://services.mbi.ucla.edu/anisotropy/>) indicated that diffraction along the b* axis was weaker than in the other two directions; therefore, reflections were subjected to a mild anisotropic truncation with resolution limits of 3.3, 3.4 and 3.3 Å along a*, b* and c*, respectively, before using them in the refinement.

Experimental phasing. Initial attempts to find a molecular replacement solution using previous class A GPCR structures as search models in Phaser⁵¹ did not generate any reliable solutions. Therefore, experimental phasing for the BRIL–GCGR(ΔECD/ΔC) fusion construct was attempted by soaking the crystals with different heavy atoms. After screening many crystals, a SAD data set was obtained from one crystal that was soaked overnight with 100 mM tantalum bromide (Ta₆Br₁₂) cluster (Jena Bioscience). The data were collected on the 23ID-D beamline (GM/CA CAT) at the Advanced Photon Source using the peak wavelength from the tantalum L3 edge (9.880 keV). A beam size of 10 μm with 5× attenuation with 1° oscillation and 1 s exposure per frame was used. A complete 360° data set was acquired from a single crystal by collecting wedges of 30° with direct and inverse beam and translating 6 times along the crystal length to expose a fresh portion of the crystal for each wedge. The SAD data set was integrated and scaled at 4 Å resolution using HKL2000. PHENIX.AutoSol⁵² was used to search for the heavy atom sites with anomalous signals at 6 Å resolution (Supplementary Fig. 7), yielding initial electron density maps.

Structure determination and refinement. While the experimental phasing work was underway, 'mixed model' molecular replacement (MR) search templates were generated based on all known GPCR structures using PHENIX.ROSETTA⁵³. The 'mixed model' templates were then superimposed together and manually trimmed to remove structurally poorly conserved regions. Parallel MR searches with Phaser were then carried out using these GPCR models along with the high resolution BRIL structure (PDB: 1M6T) as search models on a linux cluster⁵⁴. The search template based on rhodopsin (PDB: 2Z73) produced a potential solution (TFZ = 8.5). This MR solution was validated by the experimental phasing maps (Supplementary Fig. 8), and by the appearance of density not present in the search model. The experimental SIRAS phases calculated from the heavy atom were good up to 6–7 Å. However, SIRAS phases did not improve the MR maps, and thus were not used in the final refinement.

All refinement was performed using the MR solution with rounds of REFMAC5 (ref. 55), autoBUSTER⁵⁶ (Buster v2.8.0), and PHENIX.AutoBuild⁵⁷, followed by manual examination and rebuilding of refined coordinates in COOT⁵⁸ using both 2|F_o| − |F_c| and |F_o| − |F_c| maps, as well as omit maps calculated using the Bhat procedure⁵⁹ (Supplementary Fig. 9). We state 3.4 Å as the overall effective resolution of this structure; however, data to 3.3 Å were included in refinement, which improved the R/R_{free} statistics (Supplementary Table 2).

At 3.4 Å resolution, the electron densities for the majority of residues in the GCGR 7TM structure are visible, except for residues Arg 201–Trp 215 (corresponding to ECL1), and therefore these residues were not built into the GCGR 7TM structure. Residues Gly 269–Met 276, Thr 296–Asp 299, Ile 315–Ile 317, and Phe 365–Glu 371 were built into the GCGR structure, but they contained breaks in the electron densities of the C_α backbone. Hence, other conformations are possible for these residues.

Although we do not observe density for NNC0640 in the canonical ligand binding pocket, this ligand is required to obtain diffraction quality crystals of the BRIL–GCGR(ΔECD/ΔC) construct. There are two electron density blobs outside of canonical ligand-binding pocket, one at the bottom of helix VI and VII near Lys 349^{6,40}, and the other at helix I near Trp 145^{1,43}. However, both of them are too small to accommodate NNC0640.

Energy-based conformational modelling of the GCGR–glucagon complex. Glucagon was docked into the crystal structure of the GCGR ECD (PDB: 4ERS, residues 28–123) (ref. 6) using the crystal structure of the closely related GLP1R–GLP1 complex (PDB: 3IOL)⁸ as a template. All molecular modelling and docking was performed using ICM molecular modelling software⁶⁰ (v. 3.7). The initial α-helical conformation of glucagon peptide residues 11–29 was modelled based on GLP1 residues 17–35, and soft tethers between corresponding backbone C_α atoms of glucagon and GLP1 were applied. Conformation of the glucagon peptide and the interacting side chains in the ECD binding pocket were optimized (3 independent simulations of 10⁶ steps) using ICM global optimization procedure in internal coordinates^{60,61} with improved conformational energy terms for protein and peptides⁴ and 'tether weight' = 0.1.

The model of the ECD–glucagon complex was then docked and optimized with the crystal structure of the 7TM domain, completed with all side chains and ECL1.

This flexible energy-based docking/optimization procedure involved all torsion coordinates in the regions that are not defined by crystal structures, including protein backbone in the residues 1 to 10 of glucagon, and GCGR residues in ECL1 (199–218), and linker (122–126). In addition, side chain torsion variables were set free in all glucagon residues and the following regions of the GCGR model: helix I stalk region (125–136), ECL2 region (289–310), ECL3 (368–377), as well as 31 other residues lining the 7TM binding pocket.

The following three soft harmonic restraints derived from experimental crosslinking data in GLP1R and GLP1 (ref. 12) were applied between glucagon and GCGR side chains to guide docking: F6(cb)–Gln 142^{1.40}(cd), Y10(cb)–Tyr 138^{1.36}(oh), L14(cb)–Trp 295^{ECL2}(ch2). Two intramolecular harmonic restraints were also applied to glucagon residues, T7(og1)–Y10(n) and F6(cz)–Y10(cz), to facilitate N-capped formation in glucagon, as suggested by previous comparative studies of class B peptide ligands³³. Finally, a restraint was applied between the positively charged N-terminal amino group of glucagon and the carboxyl group of Glu 362^{6.53}. The importance of the carboxyl group of Glu 362^{6.53}, which is the only negatively charged residue in the 7TM binding pocket, is supported by GCGR (Supplementary Table 5) and GLP1R (ref. 29) mutation studies (Supplementary Table 6). As the N terminus is the only basic moiety in the first 10 residues of glucagon, a potential Glu 362^{6.53} salt bridge with the glucagon N terminus is the most likely explanation for the mutation effects in Glu 362^{6.53}. A total of 164 torsion variables were systematically sampled with ICM Monte-Carlo global optimization, and 455 were locally minimized in the course of this procedure. The special 'local' sampling option was applied to the ECL1 region backbone to allow efficient optimization. Three independent runs of the global optimization procedure (10⁷ steps each) resulted in similar best energy conformations within 2.5 Å r.m.s.d. for the glucagon peptide non-hydrogen atoms.

It should be noted that in the absence of glucagon, the ECD is likely to be more flexible, sampling multiple orientations relative to the 7TM domain⁶. The model also does not attempt to infer a specific functional state of the receptor, partially because such a state is not precisely defined for the 7TM crystal structure itself. For instance, NNC0640 used to stabilize the 7TM receptor fragment is a competitive antagonist to glucagon, which may have an effect on the crystallized conformation, even though NNC0640 is absent in the final structure. The accuracy of the GCGR–glucagon model may also be limited by the weak electron density of ECL2 and the top of helix V (residues 289–310), and the assumption that glucagon binds GCGR in an N-capped conformation^{7,33}.

Construction of GCGR mutants and cell transfection. The complementary DNA (cDNA) encoding the human GCGR was originally obtained from GeneCopoeia and cloned into the expression vector pcDNA3.1/V5-His-TOPO (Invitrogen) at the HindIII and EcoRI sites. The single and double mutants were constructed by PCR-based site directed mutagenesis. CHO-K1 cells were seeded onto 96-well poly-D-lysine treated cell culture plates (PerkinElmer) at a density of 2.7×10^4 cells per well. After overnight culture, the cells were transiently transfected with wild-type or mutant GCGR DNA using Lipofectamine 2000 transfection reagent (Invitrogen).

Whole-cell glucagon binding assay. Cells were harvested 24 h after transfections, washed twice and incubated with blocking buffer (F12 supplemented with 33 mM HEPES, pH 7.4, and 0.1% bovine serum albumin (BSA)) for 2 h at 37 °C. For homogeneous binding, the cells were incubated in binding buffer with constant concentration of ¹²⁵I-glucagon (40 pM) and different concentrations of unlabelled glucagon (0.02 nM to 5 μM) at room temperature for 3 h. Cells were washed three times with ice-cold PBS and lysed by 50 μl lysis buffer (PBS supplemented with 20 mM Tris-HCl, 1% Triton X-100, pH 7.4). The plates were subsequently counted for radioactivity (counts per minute, CPM) in a scintillation counter (MicroBeta2 Plate Counter, PerkinElmer) using a scintillation cocktail (OptiPhase SuperMix, PerkinElmer). Specific binding was determined by subtracting non-specific binding observed in the presence of 5 μM unlabelled glucagon.

Expression level quantification of constructed GCGR in cells by flow cytometry. Approximately 1×10^5 transfected CHO-K1 cells were blocked with PBS containing 5% BSA at room temperature for 15 min and then incubated with 1:100 diluted primary antibody (anti-GCGR, Epitomics) at room temperature for 1 h. The cells were then washed three times with PBS containing 1% BSA followed by a 1 h incubation with anti-rabbit Alexa-488-conjugated secondary antibody (1:300, Invitrogen) at 4 °C in the dark. After washes, the cells were resuspended in 200 μl of PBS containing 1% BSA for detection in a flow cytometer (Accuri C6, BD Biosciences) using laser excitation and emission wavelengths of 488 and 519 nm, respectively. For each measurement, approximately 20,000 cellular events

were collected and fluorescence intensity of positive expression cell population calculated.

NNC0640 binding assay (cell membrane based binding). NNC0640 binding was analysed using plasma membranes prepared from HEK293T cells transiently expressing GCGR constructs. Approximately 1.2×10^8 transfected HEK293T cells were harvested, suspended in 10 ml ice-cold membrane binding buffer (25 mM Tris-HCl, 0.1% BSA and 1 mM EDTA, pH 7.4) and centrifuged for 5 min at 200g. The resulting pellet was resuspended in cold membrane binding buffer, pulled through a 25G × 1 inch needle four times and centrifuged for 5 min at 20,000g. The precipitate containing the plasma membranes was suspended in membrane binding buffer containing protease inhibitor (Sigma-Aldrich) and stored at –80 °C. Protein concentration was determined using a protein BCA assay kit (Pierce Biotechnology).

For homogeneous binding, cell membrane homogenates (20 μg protein per well) were incubated in membrane binding buffer with constant concentration of ³H-NNC0640 (50 nM, labelled by PerkinElmer) and serial dilutions of unlabelled NNC0640 (1.26 nM to 100 μM) at room temperature for 5 h. Nonspecific binding was determined in the presence of 100 μM NNC0640. Following incubation, the samples were filtered rapidly in vacuum through glass fibre filter plates (Millipore). After soaking and rinsing 4 times with ice-cold binding buffer, the filters were dried and counted for radioactivity in a scintillation counter (PerkinElmer).

Western blot. Protein samples were prepared as above, separated by 10% SDS-PAGE and transferred to nitrocellulose membranes. After a 2 h incubation with blocking buffer, the membranes were incubated with 1:1,000 primary antibody (anti-V5, Sigma) overnight. The membranes were then washed three times with TBS-T buffer (0.05 M Tris, 0.15 M NaCl, 0.1% (v/v) Tween) followed by a 2 h incubation with anti-mouse horseradish peroxidase-conjugated secondary antibody (1:1,000, Cell Signaling Technology). The membranes were washed again and then detected with SuperSignal West Dura Substrate (ThermoScientific) according to the manufacturer's instructions. Each membrane was exposed to X-ray film for detecting the blots. Bands were quantified with Quantity One Software (Bio-Rad).

Statistical analysis. Results are presented as means ± s.e.m. Changes in specific radiolabelled ligands binding and cell surface expression of GCGR constructs were normalized to those measured with wild-type GCGR control (100%). IC₅₀ values in binding assay were determined by nonlinear regression analysis using the Prism 5 software (GraphPad Software).

45. Hanson, M. A. *et al.* Profiling of membrane protein variants in a baculovirus system by coupling cell-surface detection with small-scale parallel expression. *Protein Expr. Purif.* **56**, 85–92 (2007).
46. Misquitta, Y. & Caffrey, M. Detergents destabilize the cubic phase of monoolein: implications for membrane protein crystallization. *Biophys. J.* **85**, 3084–3096 (2003).
47. Cherezov, V., Peddi, A., Muthusubramaniam, L., Zheng, Y. F. & Caffrey, M. A robotic system for crystallizing membrane and soluble proteins in lipidic mesophases. *Acta Crystallogr. D* **60**, 1795–1807 (2004).
48. Xu, F., Liu, W., Hanson, M. A., Stevens, R. C. & Cherezov, V. Development of an automated high throughput LCP-FRAP assay to guide membrane protein crystallization in lipid mesophases. *Cryst. Growth Des.* **11**, 1193–1201 (2011).
49. Cherezov, V. *et al.* Rastering strategy for screening and centring of microcrystal samples of human membrane proteins with a sub-10 μm size X-ray synchrotron beam. *J. R. Soc. Interface* **6** (Suppl 5), S587–S597 (2009).
50. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
51. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
52. Terwilliger, T. C. *et al.* Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D* **65**, 582–601 (2009).
53. Terwilliger, T. C. *et al.* phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. *J. Struct. Funct. Genomics* **13**, 81–90 (2012).
54. Schwarzenbacher, R., Godzik, A. & Jaroszewski, L. The JCSG MR pipeline: optimized alignments, multiple models and parallel searches. *Acta Crystallogr. D* **64**, 133–140 (2008).
55. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
56. BUSTER v. 2.8.0 (Global Phasing, Cambridge, UK 2009).
57. Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D* **64**, 61–69 (2008).
58. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
59. Bhat, T. Calculation of an OMIT map. *J. Appl. Crystallogr.* **21**, 279–281 (1988).
60. ICM Manual v. 3.0 (MolSoft, La Jolla, California 2012).
61. Arnautova, Y. A., Abagyan, R. A. & Totrov, M. Development of a new physics-based internal coordinate mechanics force field and its application to protein loop modeling. *Proteins* **79**, 477–498 (2011).

Ptpn11 deletion in a novel progenitor causes metachondromatosis by inducing hedgehog signalling

Wentian Yang¹, Jianguo Wang¹, Douglas C. Moore¹, Haipei Liang¹, Mark Dooner², Qian Wu³, Richard Terek¹, Qian Chen¹, Michael G. Ehrlich¹, Peter J. Quesenberry² & Benjamin G. Neel⁴

The tyrosine phosphatase SHP2, encoded by *PTPN11*, is required for the survival, proliferation and differentiation of various cell types^{1,2}. Germline activating mutations in *PTPN11* cause Noonan syndrome, whereas somatic *PTPN11* mutations cause childhood myeloproliferative disease and contribute to some solid tumours. Recently, heterozygous inactivating mutations in *PTPN11* were found in metachondromatosis, a rare inherited disorder featuring multiple exostoses, enchondromas, joint destruction and bony deformities^{3,4}. The detailed pathogenesis of this disorder has remained unclear. Here we use a conditional knockout (floxed) *Ptpn11* allele (*Ptpn11^{fl/fl}*) and Cre recombinase transgenic mice to delete *Ptpn11* specifically in monocytes, macrophages and osteoclasts (lysozyme M-Cre; LysMCre) or in cathepsin K (Ctsk)-expressing cells, previously thought to be osteoclasts. *LysMCre;Ptpn11^{fl/fl}* mice had mild osteopetrosis. Notably, however, *CtskCre;Ptpn11^{fl/fl}* mice developed features very similar to metachondromatosis. Lineage tracing revealed a novel population of CtskCre-expressing cells in the perichondrial groove of Ranvier that display markers and functional properties consistent with mesenchymal progenitors. Chondroid neoplasms arise from these cells and show decreased extracellular signal-regulated kinase (ERK) pathway activation, increased Indian hedgehog (*Ihh*) and parathyroid hormone-related protein (*Pthrp*, also known as *Pthlh*) expression and excessive proliferation. Shp2-deficient chondroprogenitors had decreased fibroblast growth factor-evoked

ERK activation and enhanced *Ihh* and *Pthrp* expression, whereas fibroblast growth factor receptor (FGFR) or mitogen-activated protein kinase kinase (MEK) inhibitor treatment of chondroid cells increased *Ihh* and *Pthrp* expression. Importantly, smoothened inhibitor treatment ameliorated metachondromatosis features in *CtskCre;Ptpn11^{fl/fl}* mice. Thus, in contrast to its pro-oncogenic role in haematopoietic and epithelial cells, *Ptpn11* is a tumour suppressor in cartilage, acting through a FGFR/MEK/ERK-dependent pathway in a novel progenitor cell population to prevent excessive *Ihh* production.

Cartilage tumours, including exostoses, enchondromas and chondrosarcomas, comprise ~20% of skeletal neoplasms⁵. Benign and malignant cartilaginous tumours can arise sporadically, but cartilage tumour syndromes, including hereditary multiple exostoses, the multiple enchondromatosis disorders (Ollier disease and Maffucci syndrome) and metachondromatosis, also exist^{6,7}. The cellular and molecular pathogenesis of most cartilage tumours is incompletely understood.

Metachondromatosis is an autosomal dominant tumour syndrome featuring multiple exostoses and enchondromas^{6,7}. Recently, heterozygous early frameshift or nonsense mutations in *PTPN11* were identified in >50% of metachondromatosis cases^{3,4}. *PTPN11* encodes the non-receptor protein tyrosine phosphatase SHP2, which is required for RAS/ERK pathway activation in most receptor tyrosine kinase, cytokine receptor, and integrin signalling pathways^{1,2}. Germline activating

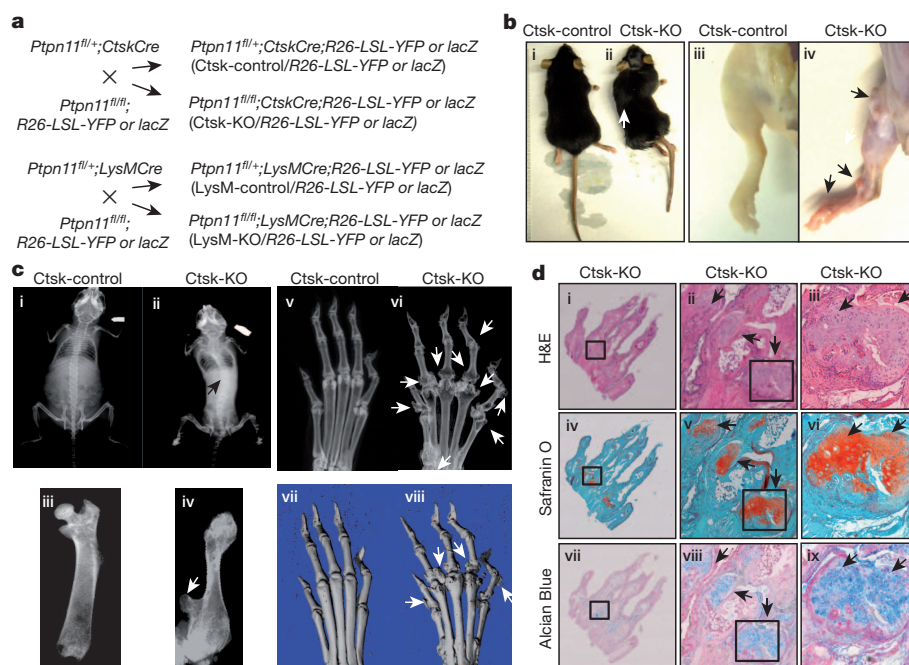


Figure 1 | *Ptpn11* deletion in Ctsk-expressing cells causes metachondromatosis. **a**, Schemes for generating Ctsk-KO, LysM-KO and control mice. **b**, **c**, Gross images (**b**) and Faxitron/μ-CT radiographs (**c**) of 12-week-old Ctsk-KO mice showing dwarfism and scoliosis (**b** ii, white arrow; **c** ii, black arrow), increased bone mineral density (**c** ii, iv; arrows) and multiple exostoses of knees, ankles and metatarsals (**b** iv and **c** ii, iv, vi, viii; arrows) with joint destruction. **d**, Sagittal sections of metatarsal joints stained with haematoxylin and eosin (H&E; i-iii), Safranin O (iv-vi) and Alcian Blue (vii-ix) showing cartilaginous exostoses and enchondromas (arrows) in Ctsk-KO mice. Images in i, iv, vii are magnified ×2, in ii, v, viii ×10 and in iii, vi, ix ×20. Images in ii, v, viii and iii, vi, ix are magnified (×10) views of boxed areas in i, iv, vii and ii, v, viii, respectively. Data shown are representative images; each analysis was performed on at least five mice per genotype.

¹Department of Orthopaedics, Brown University Alpert Medical School and Rhode Island Hospital, Providence, Rhode Island 02903, USA. ²Department of Medicine and COBRE Center for Stem Cell Biology, Rhode Island Hospital and Brown University Alpert Medical School, Providence, Rhode Island 02903, USA. ³Department of Pathology and Laboratory Medicine, University of Connecticut Health Center, Farmington, Connecticut 06030, USA. ⁴Princess Margaret Cancer Center, University Health Network, and Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 1L7, Canada.

mutations in *PTPN11* cause Noonan syndrome, whereas mutations that impair SHP2 catalytic activity cause LEOPARD syndrome (an acronym for multiple lentigenes, ECG conduction abnormalities, ocular hypertelorism, pulmonic stenosis, abnormal genitalia, retardation of growth and sensorineural deafness), both of which can feature skeletal abnormalities⁸. Somatic activating mutations in *PTPN11* are the most common cause of juvenile myelomonocytic leukaemia and contribute to other leukaemias and some solid tumours^{1,2}. Although *PTPN11* is a well-established human oncogene, it is unclear how heterozygous loss-of-function *PTPN11* alleles cause cartilage neoplasms.

Global *Ptpn11* deletion results in early embryonic lethality^{9,10}, whereas postnatal deletion has context-dependent effects^{1,2}. To assess the role of Shp2 in osteoclasts, we crossed *Ptpn11*^{fl/fl} mice¹⁰ to mice expressing Cre under the control of the endogenous *LysM*¹¹ or *Ctsk*¹² promoter. The *LysM* promoter is active in monocytes, macrophages and osteoclast precursors¹¹, whereas the *Ctsk* promoter reportedly is active only in mature osteoclasts^{12,13}. These crosses generated *Ptpn11*^{fl/+}; *LysMCre* and *Ptpn11*^{fl/fl}; *LysMCre* (hereafter, *LysM*-control and *LysM*-knockout (KO)) and *Ptpn11*^{fl/+}; *CtskCre* and *Ptpn11*^{fl/fl}; *CtskCre* (hereafter, *Ctsk*-control and *Ctsk*-KO) mice, respectively (Fig. 1a).

Neither *LysM*-control nor *Ctsk*-control mice had a discernible phenotype, so we focused all subsequent analyses on *LysM*-KO and *Ctsk*-KO mice. Shp2 levels were reduced by >80% in bone-marrow-derived macrophages and osteoclasts in *LysM*-KO and *Ctsk*-KO mice (Supplementary Fig. 1a and data not shown). *LysM*-KO and *Ctsk*-KO mice were born at the expected Mendelian ratios and appeared to be normal for the first 3 weeks after birth. Subsequently, *LysM*-KO mice developed mild, age-related osteopetrosis (Supplementary Fig. 1b and data not shown). By contrast, within 8 weeks after birth, *Ctsk*-KO mice exhibited a dramatic skeletal phenotype, comprising decreased body length, increased bone mineral density, scoliosis, metaphyseal exostoses and markedly decreased mobility (Fig. 1b–d and Supplementary Video 1).

Sections of hindlimb paw and knee joints from 12-week-old *Ctsk*-KO mice revealed multiple exostoses and enchondromas at the metaphyses of their metatarsals and phalanges (Fig. 1d), tibiae and femurs (Supplementary Fig. 1c, d), and other bones (data not shown), features reminiscent of metachondromatosis. As heterozygous *PTPN11* frameshift mutations cause metachondromatosis^{3,4}, these findings indicate that *PTPN11* is a cartilage tumour-suppressor gene, and suggest that loss (or silencing) of the remaining *PTPN11* allele is required for tumour formation.

To identify the cells responsible for metachondromatosis-like disease in *Ctsk*-KO mice, we first injected bone marrow from 6-week-old *Ctsk*-KO and *Ctsk*-control mice (C57/BL6; CD45.2) into lethally irradiated 3-week-old recipients (B6.SJL; CD45.1). Recipient mice exhibited high chimaerism (Supplementary Fig. 2a, b), but did not develop cartilage tumours in over 12 months of observation. Consistent with the osteopetrosis seen in *LysM*-KO mice, recipients had increased bone mineral density (Supplementary Fig. 2c). Clearly, however, cartilage tumours in *Ctsk*-KO mice are not due to altered osteoclast development or function.

Next, we performed lineage-tracing studies using *Rosa26-lox-stop-lox-lacZ* (R26-LSL-lacZ) or *Rosa26-LSL-YFP* (R26-LSL-YFP) Cre reporter mice. Notably, *CtskCre*, but not *LysMCre*, was expressed in a subset of perichondrial cells within the so-called groove of Ranvier (Fig. 2a). Sections from knee joints collected at postnatal day (P)10 revealed expansion of a cluster of Alcian Blue/Safranin O-positive cells in this region in *Ctsk*-KO mice, but not in controls (Fig. 2b, boxed region and Supplementary Fig. 1c). By postnatal week 2, the yellow fluorescent protein (YFP)⁺ cell population had increased and differentiated into ectopic cartilaginous tissue in compound *Ctsk*-KO;YFP reporter mice (Fig. 2c, boxed region). Exostoses were palpable at 6 weeks and visible by 8–12 weeks. In compound *Ctsk*-KO;YFP reporter mice, these lesions consisted of YFP⁺ chondroid cells at various stages of development, including proliferating, pre-hypertrophic and hypertrophic

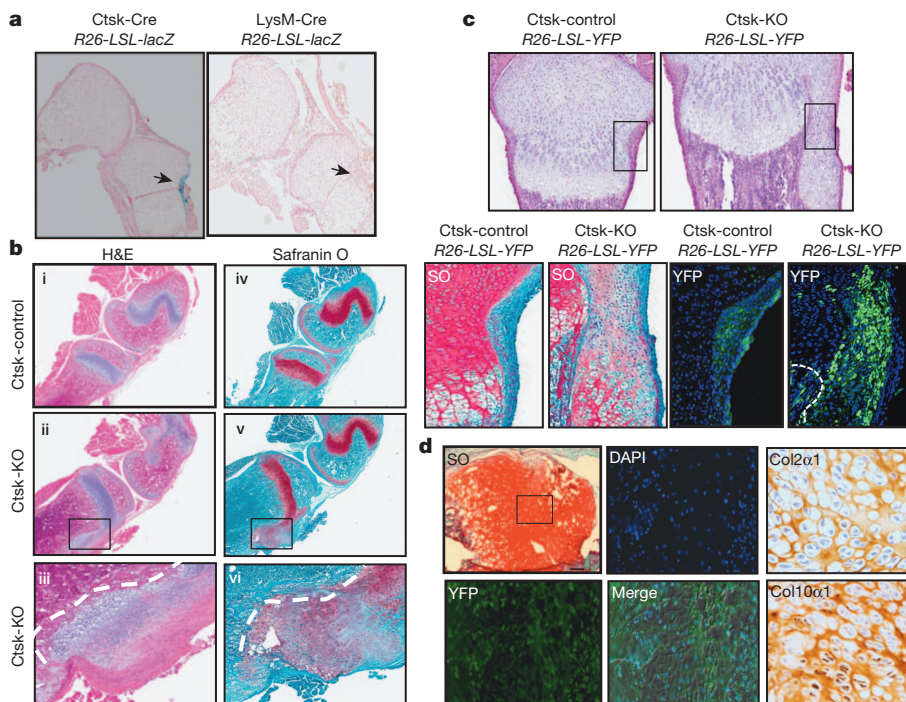


Figure 2 | Skeletal tumours in *Ctsk*-KO mice originate from perichondrial groove of Ranvier cells. **a**, 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-gal) staining of knee joint sections from 1-week-old *R26-LSL-lacZ*; *CtskCre* and *R26-LSL-lacZ*; *LysMCre* reporter mice showing that the *Ctsk* (but not the *LysM*) promoter is active not only in osteoclasts, but also in a subset of cells from the perichondrial groove of Ranvier (arrows). Images are magnified $\times 4$. **b**, Haematoxylin and eosin and Safranin O staining of knee joint sections from P10 *Ctsk*-control (i, iv) and *Ctsk*-KO (ii, iii, v, vi) mice showing expansion of cells within the perichondrial groove of Ranvier region in *Ctsk*-KO mice. Images i, ii, iv, v are magnified $\times 4$. Images in iii and vi are magnified ($\times 10$) views of boxed areas in ii and v, respectively. **c**, Haematoxylin and eosin- and Safranin O (SO)-stained sections (magnified $\times 2$) showing expanding YFP⁺ population within the perichondrial groove of Ranvier (boxed region in top panels, magnified $\times 10$ below) that also stains with Safranin O, indicative of cartilage. Dashed line marks boundary between marrow/growth plate and perichondrial groove. **d**, Frozen section of an exostosis from the metatarsal joint of *Ctsk*-KO;YFP mice showing co-localization of YFP reporter with cartilaginous tumour cells (boxed area). SO panel, $\times 4$ magnification; all other panels, $\times 10$ magnification. Note that the lesion is enriched in proliferating and pre-hypertrophic chondrocytes, as shown by overlapping Col2 α 1 and Col10 α 1 immunostaining. Each panel is a representative image from one mouse; each analysis was performed on at least three mice per genotype. DAPI, 4',6-diamidino-2-phenylindole.

chondrocytes, as revealed by cell morphology and Col2 α 1 and Col10 α 1 immunostaining (Fig. 2c, d and data not shown). Notably, nearly all chondroid tumour cells were YFP⁺ (Fig. 2c and Supplementary Fig. 2d). Hence, cartilaginous tumours in Ctsk-KO mice (and, by analogy, most likely in metachondromatosis) result from cell-autonomous lack of Shp2 in Ctsk⁺ cells from the perichondrial groove of Ranvier.

The perichondrial groove of Ranvier is believed to contain chondroprogenitors responsible for circumferential cartilage growth, but these cells are not well-characterized^{14,15}. We used flow cytometry to analyse epiphyseal cartilage cells collected from the distal femurs and proximal tibiae of Ctsk-control;YFP and Ctsk-KO;YFP mice at P10–12. Compared with controls, the frequency of YFP⁺ cartilage cells in Ctsk-KO;YFP mice was increased by ~fivefold (Fig. 3a). Within the YFP⁺ cell population, the percentage of cells expressing CD44, CD90 and CD166 (mesenchymal progenitor markers), but not CD31 (endothelial cell marker), also was increased (Fig. 3b). Staining for the stromal cell antigen Stro-1 in addition to jagged 1, markers associated with presumptive chondroprogenitors in the groove on the basis of BrdU label retention studies¹⁶, was more intense in Ctsk-KO mice (Fig. 3c). Moreover, YFP⁺ cells were capable of multi-lineage differentiation *in vitro*, as assessed by Alcian Blue, Oil Red O and Alizarin Red staining, respectively (Fig. 3d). These data indicate that Shp2 regulates the proliferation of a novel cartilage cell population characterized by Ctsk expression, which we hereafter term Ctsk⁺ chondroid progenitors (CCPs).

Multiple pathways control cartilage development and homeostasis¹⁷. IHH and PTHRP signalling are particularly important, and aberrant regulation of these pathways causes developmental defects and skeletal tumours^{18,19}. We examined chondrogenic gene expression in cartilage tumours from Ctsk-KO mice by quantitative reverse-transcription PCR (qRT-PCR). Consistent with our immunostaining data (Fig. 2d), Col2a1 and Col10a1 transcripts were increased. Furthermore, *Ihh* and *Pthrp* levels were increased substantially (Fig. 4a and Supplementary Fig. 3a).

These findings prompted us to ask whether Shp2 regulates *Ihh* and *Pthrp* production, and if so, how. During development, cells within the perichondrium make Fgf18, which can signal to adjacent cells via Fgfr3 to suppress *Ihh* expression^{20,21}. As Shp2 is required for Fgfr3 signalling in other cell types^{1,2}, we suspected that Shp2 might be required for Fgfr3-induced suppression of *Ihh* expression. We therefore examined the status of Fgfr3 signalling components and *Ihh* expression in CCPs. Erk activation, as assayed by Tyr 204 Thr 202 phosphorylation, was compromised in the absence of Shp2, whereas Akt (phospho (p)-Ser 473) and Stat1/3 (p-Tyr 807) activation were unaffected (Fig. 4a, Supplementary Fig. 3b and data not shown). Furthermore, consistent with our qRT-PCR data, *Ihh* messenger RNA and protein were increased in Shp2-deficient CCPs (Fig. 4a). *Ihh* antibody specificity was confirmed by immunostaining of growth plate cartilage (Supplementary Fig. 3c).

CCPs are rare, rendering their detailed biochemical analysis unfeasible. We therefore tested the effects of Shp2 depletion in ATDC5 chondroid cells by stably expressing either of two short hairpin RNAs (shRNAs) targeting mouse *Ptpn11*. As in Ctsk-KO mice (Fig. 4a), Fgf18-evoked Erk activation was decreased, whereas *Ihh* and *Pthrp* levels were increased in Shp2-deficient cells (Fig. 4b). Conversely, FGFR (PD173074) or MEK (UO126) inhibition led to enhanced *Ihh* and *Pthrp* expression in parental ATDC5 cells (Fig. 4c).

Ihh signalling evokes *Pthrp* production²². Our data, in addition to previous studies²³, suggested that increased *Ihh* levels might be pathogenic in metachondromatosis. If so, then blocking or attenuating *Ihh* signalling might slow and/or prevent the disease. To test this hypothesis, control (wild-type) and Ctsk-KO mice (9 per group) were gavaged daily with the smoothened inhibitor PF-04449913 (SMOi, 100 μ g g⁻¹ body weight) or vehicle control (0.5% methylcellulose), beginning at 5 weeks of age (when early lesions were present) and continuing for the succeeding 4 weeks. Skeletal phenotype was assessed by X-ray, micro-computed tomography (μ -CT) and histology. Notably, SMOi treatment significantly reduced the number of exostoses in Ctsk-KO

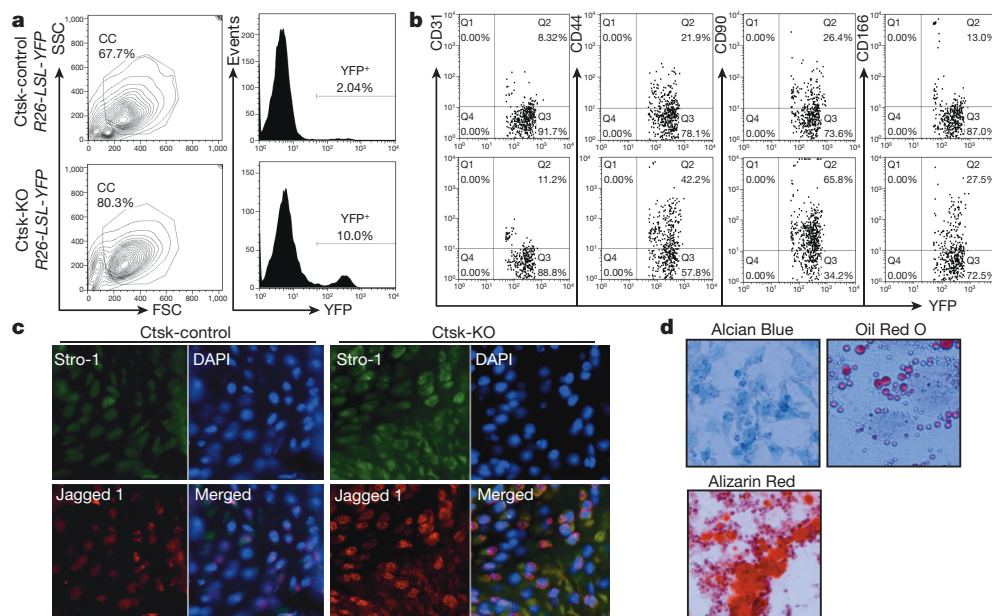


Figure 3 | *Ptpn11* deletion in Ctsk-expressing cells causes expansion of novel chondroprogenitor cell population within the perichondrial groove of Ranvier. **a**, Flow cytometric analysis showing YFP⁺ cells from pooled epiphyseal cartilage from 5–7 Ctsk-control;YFP mice; note increased percentage of such cells in 2-week-old Ctsk-KO;YFP mice. CC, chondroid cells. **b**, Flow cytometric analysis of YFP⁺ perichondrial cells showing staining for CD31, CD44, CD90 and CD166. Data in panels **a** and **b** are from a single experiment; similar results were obtained in two additional experiments. **c**, Immunofluorescence micrograph (magnified $\times 20$) showing Stro-1 and

jagged 1 expression in YFP⁺ perichondrial cells. Nuclei are stained with DAPI. Note enhanced intensity of Stro-1 and jagged 1 staining in Ctsk-KO cells. Data shown are from single mice of each genotype; two additional mice were analysed for each genotype with similar results. **d**, CCPs give rise to cartilage, fat and bone. Fluorescent-activated cell sorting (FACS)-purified YFP⁺ cells from 5–7 mice were subjected to differentiation assays in triplicate. After 2–3 weeks of culture (see Methods), cells were fixed and stained with Alcian Blue, Oil Red O and Alizarin Red to visualize the formation of cartilage, fat and bone tissue (magnified $\times 20$), respectively.

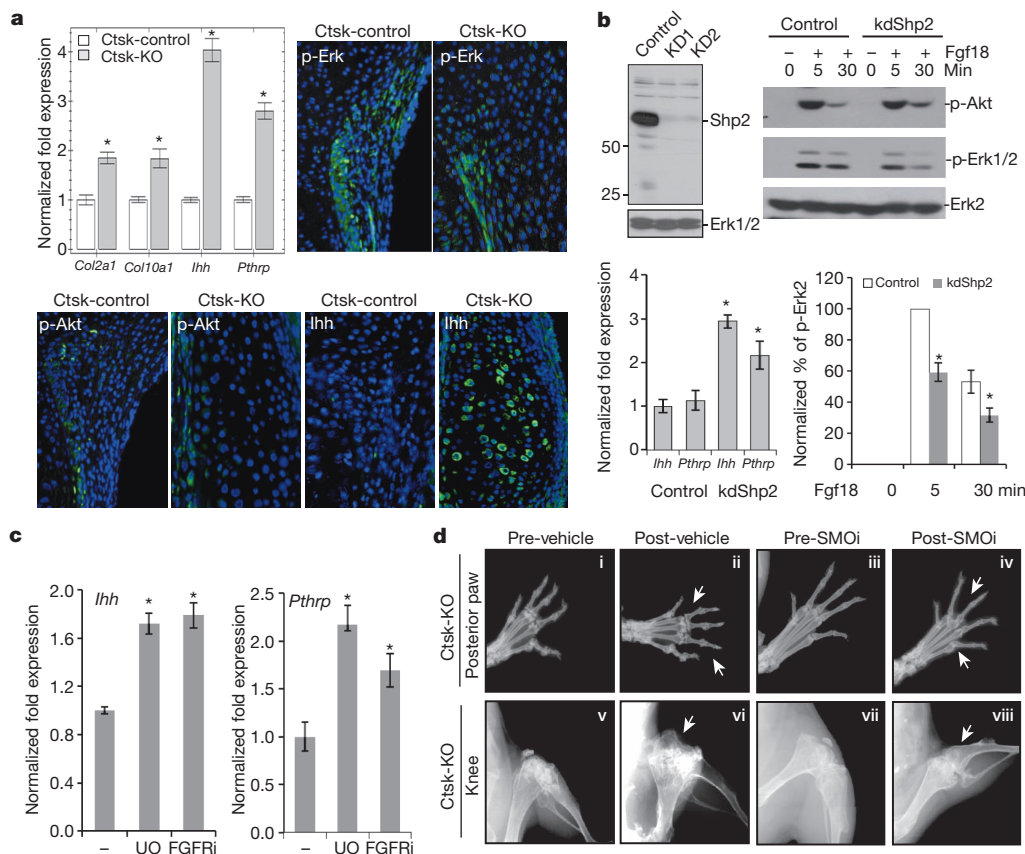


Figure 4 | Shp2 deficiency impairs Erk activation but promotes *Ihh* and *Pthrp* expression. **a**, Top left, qRT-PCR showing increased *Col2a1*, *Col10a1*, *Ihh* and *Pthrp* expression in laser-captured cartilaginous cells from exostoses in four mice per genotype, compared with normal articular cartilage cells (mean \pm s.d.; * P < 0.05, two-tailed Student's t test). Top right, immunostaining of representative paraffin sections from perichondrial groove of Ctsk-KO and control mice. Note the decreased number of p-Erk⁺ cells (75.4% in Ctsk-control versus 32.2% in Ctsk-KO; n = 3 mice). Bottom, note the increased *Ihh* expression but unchanged p-Akt staining in Ctsk-KO, compared with control, mice. **b**, Top left, immunoblot showing Shp2 in ATDC5 knockdown (KD) cells, stably expressing shRNAs against mouse *Ptpn11* (ATDC5-KD1, ATDC5-KD2) or scrambled control hairpin. Top right, Representative blot showing that Shp2 deficiency decreases Erk activation in response to Fgf18; data from multiple experiments (n = 3) showing p-Erk levels (compared with control at 5 min, mean \pm s.d.; P < 0.05, two-tailed Student's

mice (Fig. 4d and Supplementary Figs 4–7) and markedly improved their mobility (Supplementary Video 1), without apparent effects on overall growth rate (Supplementary Fig. 8). Importantly, SMOi levels in treated mice were adequate to suppress *Ihh* target gene expression in exostoses (Supplementary Fig. 7b).

Our findings strongly suggest that metachondromatosis results from loss of SHP2, specifically in CCPs, a heretofore poorly characterized population within the perichondrial groove of Ranvier, which is believed to function as a stem-cell niche for joints¹⁶ and a reservoir for the germinal layer cells of the growth plate²⁴. Cells within the groove of Ranvier express high levels of FGFR3 (ref. 25), and their removal prevents longitudinal bone growth²⁶. Emerging evidence shows that groove of Ranvier cells can migrate into articular cartilage¹⁶, implicating them in maintaining cartilage homeostasis and possibly in degenerative joint diseases, such as osteoarthritis. Indeed, in lineage-tracing studies of normal mice, we noticed YFP⁺ cells migrating towards articular cartilage (Supplementary Fig. 9, arrows, and data not shown). On the basis of our mouse metachondromatosis model, we propose that SHP2, acting downstream of FGFR3 and upstream of the RAS/ERK pathway, regulates CCP proliferation and chondrogenic

differentiation. Consequently, *PTPN11* deficiency in these cells promotes excessive proliferation, chondrogenic differentiation and cartilage tumours. Metachondromatosis is associated with heterozygous inactivating mutations in *PTPN11*, yet Ctsk-control mice are normal, whereas Ctsk-KO mice exhibit metachondromatosis-like features. Although *PTPN11* gene dosage effects could differ in mouse and man (and thus 50% reduction in SHP2 level might cause metachondromatosis in humans but not in mice), we think it is more likely that loss of the remaining *PTPN11* allele (for example, by loss of heterozygosity or silencing) is required to cause cartilage tumours in metachondromatosis. If so, then unlike its oncogenic role in juvenile myelomonocytic leukaemia, other hematologic malignancies and solid tumours^{1,2}, *PTPN11* is a tumour suppressor in cartilage. Liver-specific *Ptpn11* deletion reportedly results in hepatocellular carcinoma²⁷. However, we have not seen liver tumours in our *Ptpn11* conditional knockout mice crossed to the same Cre line (F.H. & B.G.N., manuscript in preparation), nor is *PTPN11* mutated in human hepatocellular carcinoma. Moreover, our biochemical and pharmacological analysis, together with previous studies, provide a parsimonious and attractive explanation for the

apparently paradoxical pro- and anti-oncogenic effects of *PTPN11*. In both cases, SHP2 is a critical regulator of ERK. The activating *PTPN11* mutations associated with cancer promote proliferation and survival, at least in part via increased ERK activation. Similarly, overexpression or increased activation of normal SHP2 binding proteins such as GAB2, or the presence of pathologic SHP2 binding proteins such as *Helicobacter pylori* CagA²⁸, can hyperactivate ERK and contribute to various malignancies. Conversely, SHP2 deficiency is oncogenic in CCPs because in these cells, ERK normally represses the expression of the growth stimulator IHH (which, in turn, stimulates PTHRP production). Future studies should focus on better defining the properties of CCPs, determining whether *PTPN11* also acts as a tumour suppressor in other cartilage neoplasms, including chondrosarcoma, and most importantly, on testing the effects of smoothened inhibition in metachondromatosis patients. Finally, given our proposed mechanism of metachondromatosis pathogenesis, our results call for caution in the long-term use of MEK or ERK inhibitors.

METHODS SUMMARY

Ptpn11 floxed (*Ptpn11*^{fl})¹⁰, *cathepsin K-Cre* (*Ctsk-Cre*)¹², *Rosa26-LSL-lacZ* (*R26-LSL-lacZ*)²⁹, and *Rosa26-LSL-YFP* (*R26-LSL-YFP*)³⁰ Cre reporter mice were on C57BL/6 background. PCR genotyping was performed as described^{10,12,29,30}; conditions are available from W.Y. Animal studies were approved by the Institutional Animal Care and Use Committee at Rhode Island Hospital. Mice of both genders were used for this study. Antibodies and reagents are detailed in Methods. CCPs isolated from epiphyseal cartilage of 2-week-old *Ctsk*-control;YFP and *Ctsk*-KO;YFP mice (detailed in Methods) were analysed by flow cytometry or subjected to multi-lineage differentiation assays. For flow cytometry, cells were stained with fluorescence-labelled antibodies, and analysed on a BD LSR II flow cytometer using FlowJo software (TreeStar). YFP⁺ cells were purified by using a BD Influx cell sorter (BD Bioscience). ATDC5 cells stably expressing mouse *Ptpn11* shRNAs or a scrambled control were generated in pSuper(retro)/puro (Oligoengine). RNA was extracted from cells or laser-dissected lesions using RNeasy (Qiagen), complementary DNA was synthesized using iScript (Bio-Rad), and qRT-PCR was performed with the iQ SYBR Green kit. Values were normalized to *Gapdh* levels, and are expressed as fold-change over control. Primer sequences are available from W.Y.. Femurs, tibiae and paws were fixed in 4% paraformaldehyde, decalcified in 0.5 M EDTA, and sections (5 µm) were stained with haematoxylin and eosin, Alcian Blue or Safranin O. Immunohistochemistry was performed using peroxidase-coupled secondary antibodies, with diaminobenzidine as the substrate. X-gal staining was performed as described¹². Digital imaging was performed using Faxitron (plain films) and desktop microcomputed tomographic (µ-CT40, Scanco Medical) systems. The number of exostoses was quantified from these images. Immunoblotting and detection by enhanced chemiluminescence (Amersham) were performed as described¹⁰. Differences between groups were evaluated by Student's *t* test, with *P* < 0.05 considered significant.

Full Methods and any associated references are available in the online version of the paper.

Received 3 May 2012; accepted 21 June 2013.

Published online 17 July 2013.

1. Neel, B. G., Chan, G. & Dhanji, S. in *Handbook of Cell Signaling* Vol. 2 (eds Bradshaw, R. A. & Dennis, E. A.) Ch. 98, 771–809 (2009).
2. Grossmann, K. S., Rosario, M., Birchmeier, C. & Birchmeier, W. The tyrosine phosphatase Shp2 in development and cancer. *Adv. Cancer Res.* **106**, 53–89 (2010).
3. Bowen, M. E. *et al.* Loss-of-function mutations in *PTPN11* cause metachondromatosis, but not Ollier disease or Maffucci syndrome. *PLoS Genet.* **7**, e1002050 (2011).
4. Sobreira, N. L. *et al.* Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. *PLoS Genet.* **6**, e1000991 (2010).
5. Bovee, J. V., Hogendoorn, P. C., Wunder, J. S. & Alman, B. A. Cartilage tumours and bone development: molecular pathology and possible therapeutic targets. *Nature Rev. Cancer* **10**, 481–488 (2010).
6. Pannier, S. & Legeai-Mallet, L. Hereditary multiple exostoses and enchondromatosis. *Best Pract. Res. Clin. Rheumatol.* **22**, 45–54 (2008).
7. Pansuriya, T. C., Kroon, H. M. & Bovee, J. V. Enchondromatosis: insights on the different subtypes. *Int. J. Clin. Exp. Pathol.* **3**, 557–569 (2010).

8. Tartaglia, M., Gelb, B. D. & Zenker, M. Noonan syndrome and clinically related disorders. *Best Pract. Res. Clin. Endocrinol. Metab.* **25**, 161–179 (2011).
9. Saxton, T. M. *et al.* Abnormal mesoderm patterning in mouse embryos mutant for the SH2 tyrosine phosphatase Shp-2. *EMBO J.* **16**, 2352–2364 (1997).
10. Yang, W. *et al.* An Shp2/SFK/Ras/Erk signaling pathway controls trophoblast stem cell survival. *Dev. Cell* **10**, 317–327 (2006).
11. Clausen, B. E., Burkhardt, C., Reith, W., Renkawitz, R. & Forster, I. Conditional gene targeting in macrophages and granulocytes using LysMcre mice. *Transgenic Res.* **8**, 265–277 (1999).
12. Nakamura, T. *et al.* Estrogen prevents bone loss via estrogen receptor α and induction of Fas ligand in osteoclasts. *Cell* **130**, 811–823 (2007).
13. Dodds, R. A., Connor, J. R., Drake, F., Feild, J. & Gowen, M. Cathepsin K mRNA detection is restricted to osteoclasts during fetal mouse development. *J. Bone Miner. Res.* **13**, 673–682 (1998).
14. Langenskiöld, A. Role of the ossification groove of Ranvier in normal and pathologic bone growth: a review. *J. Pediatr. Orthop.* **18**, 173–177 (1998).
15. Shapiro, F., Holtrop, M. E. & Glimcher, M. J. Organization and cellular biology of the perichondrial ossification groove of ranvier: a morphological study in rabbits. *J. Bone Joint Surg. Am.* **59**, 703–723 (1977).
16. Karlsson, C., Thormemo, M., Henriksson, H. B. & Lindahl, A. Identification of a stem cell niche in the zone of Ranvier within the knee joint. *J. Anat.* **215**, 355–363 (2009).
17. Goldring, M. B., Tsuchimochi, K. & Ijiri, K. The control of chondrogenesis. *J. Cell. Biochem.* **97**, 33–44 (2006).
18. Hopyan, S. *et al.* A mutant PTH/PTHrP type I receptor in enchondromatosis. *Nature Genet.* **30**, 306–310 (2002).
19. Tiet, T. D. *et al.* Constitutive Hedgehog signaling in chondrosarcoma up-regulates tumor cell proliferation. *Am. J. Pathol.* **168**, 321–330 (2006).
20. Liu, Z., Xu, J., Colvin, J. S. & Ornitz, D. M. Coordination of chondrogenesis and osteogenesis by fibroblast growth factor 18. *Genes Dev.* **16**, 859–869 (2002).
21. Ohbayashi, N. *et al.* FGF18 is required for normal cell proliferation and differentiation during osteogenesis and chondrogenesis. *Genes Dev.* **16**, 870–879 (2002).
22. Kronenberg, H. M. PTHrP and skeletal development. *Ann. NY Acad. Sci.* **1068**, 1–13 (2006).
23. Murakami, S. *et al.* Constitutive activation of MEK1 in chondrocytes causes Stat1-independent achondroplasia-like dwarfism and rescues the *Fgf3*-deficient mouse phenotype. *Genes Dev.* **18**, 290–305 (2004).
24. Fenichel, I., Evron, Z. & Nevo, Z. The perichondrial ring as a reservoir for precartilaginous cells. *In vivo* model in young chicks' epiphysis. *Int. Orthop.* **30**, 353–356 (2006).
25. Robinson, D. *et al.* Fibroblast growth factor receptor-3 as a marker for precartilaginous stem cells. *Clin. Orthop. Relat. Res.* **S163**–S175 (1999).
26. Rodriguez, J. I., Delgado, E. & Paniagua, R. Changes in young rat radius following excision of the perichondrial ring. *Calcif. Tissue Int.* **37**, 677–683 (1985).
27. Bard-Chapeau, E. A. *et al.* *Ptpn11/Shp2* acts as a tumor suppressor in hepatocellular carcinogenesis. *Cancer Cell* **19**, 629–639 (2011).
28. Hatakeyama, M. Oncogenic mechanisms of the *Helicobacter pylori* CagA protein. *Nature Rev. Cancer* **4**, 688–694 (2004).
29. Soriano, P. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21**, 70–71 (1999).
30. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of *EYFP* and *ECFP* into the ROSA26 locus. *BMC Dev. Biol.* **1**, 4 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Kato for *Ctsk*-Cre mice, A. Craft for review of the manuscript, X. Wang and P. Monfils for help with histology and J. Cao for helping with µ-CT analysis. This publication was made possible by the National Institutes of Health (NIH) and the National Institute for General Medicine Sciences (NIGMS) grant no. 8P20GM103468. This work was also funded by NIH R21AR57156 (to W.Y.) and R37CA49152 (to B.G.N.), the Rhode Island Hospital Orthopaedic Foundation and a grant from the Pediatric Orthopaedic Society of North America and the Orthopaedic Research and Education Foundation (to W.Y.). B.G.N. is a Canada Research Chair, Tier 1, and is also supported in part by the Ontario Ministry of Health and Long Term Care and the Princess Margaret Cancer Foundation.

Author Contributions W.Y. and B.G.N. conceived the project. J.W., D.M., H.L. and W.Y. carried out most of the experiments. M.D. and W.Y. conducted FACS sorting and analysis. H.L. performed the SMOi animal treatment experiment. J.W. performed gene expression and western blot analysis. M.D. and J.W. carried out bone marrow transplantation experiments with the advice of P.J.Q. M.D., H.L. and W.Y. performed CPC multi-lineage differentiation assays. Q.W. and R.T. performed histological staining and data interpretation. Q.C. and M.G.E. provided technical and intellectual support. W.Y. and B.G.N. analysed the data and wrote the manuscript with the help of all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.Y. (wyang@lifespan.org).

METHODS

Ptpn11 floxed (*Ptpn11^{fl}*)¹⁰, *cathepsin K-Cre* (Ctsk-Cre)¹², *Rosa26-LSL-lacZ* (*R26-LSL-lacZ*)²⁹, and *Rosa26-LSL-YFP* (*R26-LSL-YFP*)³⁰ Cre reporter mice were on C57BL/6 background.

Antibodies and reagents. The following antibodies were purchased: monoclonal anti-p-tyrosine (4G10) was from Millipore; polyclonal antibodies against p-Erk1/2, Erk2, p-Akt(Ser 473), Akt, Shp2, p-Stat1 (Tyr 701) and Stat1 were from Cell Signaling; antibodies against Ihh, Col2 α 1 and Col10 α 1 were from Santa Cruz Biotechnology and Abcam, respectively; fluorescence-labelled antibodies against CD31, CD44, CD45, CD90 and CD166 were purchased from eBioscience; and antibodies against Stro-1 and jagged 1 were purchased from Invitrogen and EpiTomics, respectively. Alexa 488-labelled goat anti-rabbit IgG and Alexa 594-labelled anti-rabbit and anti-mouse IgG were purchased from Invitrogen. Fgf18 was purchased from PeproTech. UO126 and PD173074 were from Calbiochem and Selleckbio, respectively. PF-04449913 was kindly provided by Pfizer. Alcian Blue, Alizarin Red S and Oil Red O staining solutions were purchased from Poly Scientific.

Cell isolation and culture. To isolate YFP⁺ CCPs, epiphyseal cartilage was dissected from 2-week-old Ctsk-control; YFP and Ctsk-KO; YFP mice, and digested with hyaluronidase (2.5 mg ml⁻¹, Sigma) and trypsin-EDTA (0.25%, Invitrogen) to remove soft tissues, and then with collagenase D (2.5 mg ml⁻¹, Roche) for 4–6 h to release all cartilage cells. After washing in PBS, cells were stained with fluorescence-labelled antibodies (using concentrations recommended by the manufacturers), and analysed by flow cytometry, or YFP⁺ cells were purified by FACS and placed in short-term cultures (3–4 days) in murine mesenchymal culture medium (StemCell Technologies) containing 10% FBS.

Parental ATDC5 cells were obtained from C. Phornphutkul (Brown University) and cultured in complete DMEM/F12 medium (1:1) (Invitrogen), as described³¹. shRNAs against mouse *Ptpn11* (KD1: 5'-GATTCAGAACACTGGGACTTCAAGAGTCCCCAGTGTCTGAATC-3'; KD2: 5'-GAGTAACCCTGGAGACTTCTTCAAGAGAGAAGTCTCCAGGGTTACTC-3'), or a scrambled control for KD1 (5'-TAGTACAAGTCCAAGCGGCTTCAAGAGA GCCGCTGGACTTGTACTA-3'), were introduced into the retroviral vector pSuper(retro)/puro (Oligoengine). Viral supernatants were collected from 293T cells co-transfected with each retroviral vector and Ecopac, and used to infect ATDC5 cells, which were then selected with puromycin³².

Differentiation assays. CCPs (~2 × 10⁴), purified by FACS (for YFP) from 10–14-day-old *Ctsk-R26-LSL-YFP* reporter mice, were cultured in differentiation medium for chondrocytes (DMEM with 10% FBS, 0.1 μ M dexamethasone, 0.1 mM ascorbic acid, 10 mM glycerol 2-phosphate, 1 ng ml⁻¹ TGF- β 1), adipocytes (DMEM with 10% FBS, 1 μ M dexamethasone, 0.5 mM 3-isobutyl-1-methylxanthine (IBMX), 10 μ g ml⁻¹ insulin), or osteoblasts (DMEM with 10% FBS, 0.1 μ M dexamethasone, 0.2 mM ascorbic acid, 10 mM glycerol-2-phosphate, 10 ng ml⁻¹ rhMBP2), respectively. After culturing for 2 (adipogenic or chondrogenic differentiation) or 3 weeks (osteogenesis), cells were fixed and stained with Alcian Blue, Oil Red O or Alizarin Red to visualize the formation of cartilage, fat, and bone tissue, respectively.

qRT-PCR. RNA was extracted from cultured cells or cartilage lesions enriched by laser-capture using the RNeasy kit (Qiagen). cDNA was synthesized using iScriptcDNA Synthesis Kit (Bio-Rad), and qRT-PCR was performed by using the iQ SYBR Green qPCR kit. All values were normalized to *Gapdh* levels, and qRT-PCR data were expressed as fold-increases compared with controls. Primer sequences and PCR conditions are available from W.Y. upon request.

Flow cytometry and FACS. Epiphyseal cartilage cells were stained with fluorescence-labelled antibodies, as described³³, and analysed on a BD LSR II flow cytometer. YFP⁺ cells were purified by FACS using a BD Influx cell sorter (BD Bioscience). Flow cytometric data were analysed with FlowJo software (TreeStar).

Histology. Ctsk-control and Ctsk-KO mice were euthanized at the indicated ages, and femurs, tibiae and paws were removed and fixed in 4% PFA overnight at 4 °C. Postnatal skeletal tissues were decalcified in 0.5 M EDTA before embedding. Tissue sections (5 μ m) were stained with haematoxylin and eosin, Alcian blue or Safranin O. Immunofluorescence staining was carried out using secondary antibodies conjugated to the indicated fluorophores at concentrations recommended by their manufacturers. Immunohistochemistry was performed using fluorescence- or peroxidase-coupled anti-rabbit, -mouse or -goat secondary antibodies, as per the manufacturer's instructions, with diaminobenzidine serving as the substrate. X-gal staining was performed as described¹².

Drug treatment. Two trials were performed using the smoothened inhibitor PF-04449913. In a pilot study, groups (five mice per each group) of knockout mice were treated with SMOi (100 μ g g⁻¹ body weight) or vehicle control (0.5% methylcellulose), beginning at 5 weeks of age (at which time early lesions were present) and continuing for the succeeding 4 weeks. Mice were randomized by alternate assignment to control (vehicle) or drug treatment arms. The pilot experiment showed a significant difference in number of exostoses (assessed radiographically) in the SMOi group, and led to a second study (again involving five mice each) to confirm these findings and also assess additional parameters (μ -CT, histology, gene expression). Two mice (one each from control and experimental groups, respectively) died for unknown reasons during the second trial, and were excluded from the analysis because they were removed from cages and could not be recovered. All surviving mice from both studies were included in the analyses shown in the text.

Microcomputed tomography (μ -CT) and X-ray analysis. X-ray images of the entire skeleton, knees, metatarsals and phalanges were obtained immediately after euthanasia by using a Faxitron X-ray system (Wheeling). After fixation in 4% PFA, μ -CT images of skeletal tissues were scanned with a desktop microcomputer graphic imaging system (μ CT40, Scanco Medical AG). The number of exostoses was measured from these radiographic images, as indicated in the figure legends. For these studies, mice were assigned a code number by the animal technician, and blinded quantification was carried out by W.Y.

Immunoblotting. Cells were lysed in modified NP-40 buffer (0.5% NP-40, 150 mM NaCl, 1 mM EDTA, 50 mM Tris (pH 7.4)), supplemented with a protease inhibitor cocktail (1 mM PMSF, 1 mM NaF, 1 mM sodium orthovanadate, 10 mg ml⁻¹ aprotinin, 0.5 mg ml⁻¹ antipain and 0.5 mg ml⁻¹ pepstatin), as described¹⁰. For immunoblotting, cell lysates (10–50 μ g) were resolved by SDS-PAGE, transferred to polyvinylidene difluoride membranes, and incubated with primary antibodies for 2 h or overnight at 4 °C (according to the manufacturer's instructions), followed by horseradish peroxidase-conjugated secondary antibodies. Detection was by enhanced chemiluminescence (Amersham). Signals were quantified using NIH ImageJ.

Statistical analysis. Differences between groups were evaluated by Student's *t* test. A *p* value of <0.05 was considered significant. For all of these experiments, between-group variances were similar and data were symmetrically distributed. All analyses were performed by using Excel (Microsoft) and Prism 3.0 (GraphPad).

- Hidaka, K. *et al.* Involvement of the phosphoinositide 3-kinase/protein kinase B signaling pathway in insulin/IGF-I-induced chondrogenesis of the mouse embryonal carcinoma-derived cell line ATDC5. *Int. J. Biochem. Cell Biol.* **33**, 1094–1103 (2001).
- Mohi, M. G. *et al.* Prognostic, therapeutic, and mechanistic implications of a mouse model of leukemia evoked by Shp2 (*PTPN11*) mutations. *Cancer Cell* **7**, 179–191 (2005).
- Pretzel, D. *et al.* Relative percentage and zonal distribution of mesenchymal progenitor cells in human osteoarthritic and normal cartilage. *Arthritis Res. Ther.* **13**, R64 (2011).

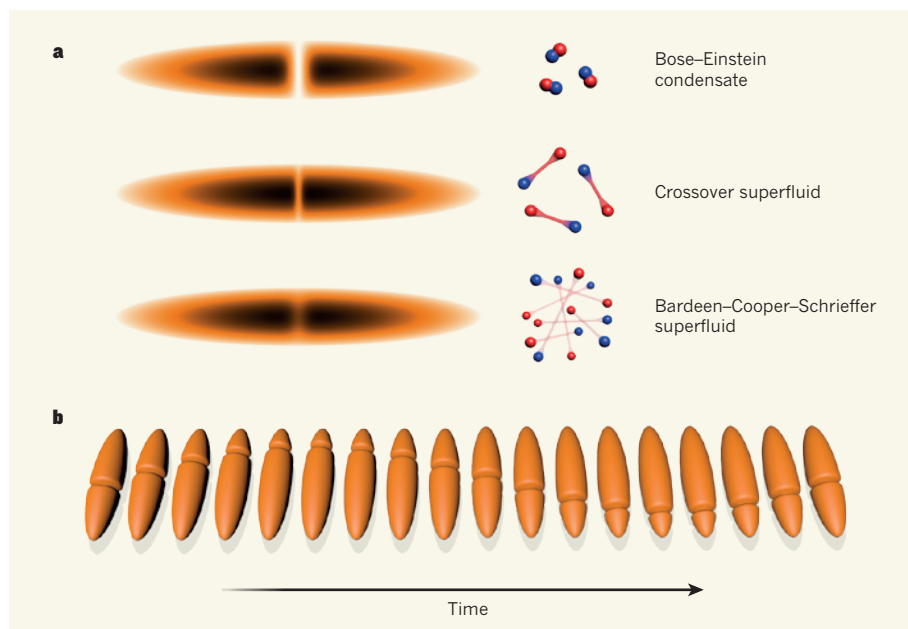


Figure 1 | Dark solitons in a fermionic superfluid. Yefsah *et al.*¹ created and observed the motion of dark solitons in an ultracold gas of strongly interacting fermionic atoms. **a**, The size of the atomic pairs (blue and red) comprising the gas determines how much such a soliton modulates the density of the gas (orange and black, with black denoting higher density and orange lower). The dark soliton (dip in the gas density) gets increasingly filled with non-condensed-gas atoms as the system moves from the Bose-Einstein condensation regime to the intermediate crossover superfluid, and then to a Bardeen-Cooper-Schrieffer superfluid. **b**, The propagation of a dark soliton in a trapped interacting Fermi gas.

arises from the existence of superfluidity in these interacting fermionic systems^{7–9}. In the crossover between the BEC and BCS regimes, the pair size is comparable to the inter-particle spacing and the gas is difficult to describe¹⁰. However, this regime is particularly appealing because its physics is the same for all different types of fermions — such a Fermi gas is said to be universal or unitary. Lacking any satisfactory theoretical description, researchers have conducted several experiments in this regime using ultracold quantum gases of fermions, and fundamental equilibrium and a few dynamical properties have been observed. One striking example was the observation of vortices, quantized units of angular momentum, which unambiguously confirmed the existence of robust superfluidity also in the unitary Fermi gas¹¹.

In their experiments, Yefsah *et al.* used the dynamical evolution of dark solitons across the BEC–BCS transition as a probe of the underlying Fermi gas. They continuously varied the pair size in an ultracold gas of lithium-6 fermionic atoms over the complete BEC–BCS crossover (Fig. 1) using a tool known as magnetic Feshbach resonance. They created dark solitons and allowed them to oscillate in the ‘harmonic trap’ they used to confine the atoms. Unexpectedly, they found that, whereas soliton oscillations deep in the BEC regime behaved quite as expected from theory, their motion was slowed down by almost a factor of 20 in the unitary and BCS regimes. The authors observed that, as these regimes are approached, the density dip that makes up a dark soliton gets

significantly more filled with non-condensed-gas atoms than expected from calculations. This filling makes the soliton heavier and thus slows down its motion^{7–9}.

Thermal and quantum fluctuations could both be responsible for these non-condensed atoms and thus for the existence of such heavy

solitons. Yefsah *et al.* have carefully analysed soliton oscillations at different temperatures and showed that the observed increase in the effective mass of the solitons is not caused by thermal fluctuations. On the other hand, existing theories underestimate the effect of quantum fluctuations, so Yefsah and colleagues’ study should be regarded as a benchmark quantum simulation to test future theories of strongly interacting Fermi gases. Regardless of whether exotic states inside the solitons, such as Andreev bound states, or completely different and as yet unknown mechanisms can generate the slow motion of the dark solitons, shedding light on this problem might turn out to be beneficial for our general understanding of strongly interacting fermions. ■

Christoph Becker is at the Institute for Laser Physics, University of Hamburg, 22761 Hamburg, Germany.
e-mail: cbecker@physnet.uni-hamburg.de

1. Yefsah, T. *et al.* *Nature* **499**, 426–430 (2013).
2. Zabusky, N. J. & Kruskal, M. D. *Phys. Rev. Lett.* **15**, 240–243 (1965).
3. Burger, S. *et al.* *Phys. Rev. Lett.* **83**, 5198–5201 (1999).
4. Becker, C. *et al.* *Nature Phys.* **4**, 496–501 (2008).
5. Weller, A. *et al.* *Phys. Rev. Lett.* **101**, 130401 (2008).
6. Stellmer, S. *et al.* *Phys. Rev. Lett.* **101**, 120406 (2008).
7. Busch, T. & Anglin, J. R. *Phys. Rev. Lett.* **84**, 2298–2301 (2000).
8. Scott, R. G., Dalfovo, F., Pitaevskii, L. P. & Stringari, S. *Phys. Rev. Lett.* **106**, 185301 (2011).
9. Spuntarelli, A., Carr, L. D., Pieri, P. & Stringari, G. C. *New J. Phys.* **13**, 035010 (2011).
10. Giorgini, S., Pitaevskii, L. P. & Stringari, S. *Rev. Mod. Phys.* **80**, 1215–1274 (2008).
11. Zwierlein, M. W., Abo-Shaeer, J. R., Schirotzek, A., Schunk, C. H. & Ketterle, W. *Nature* **435**, 1047–1051 (2005).

CELL BIOLOGY

Tumour stem cells in bone

Activation of the signalling molecule SHP2 is implicated in driving several cancers. In a newly described class of bone-progenitor cells, however, it seems that the protein acts as a tumour suppressor. SEE LETTER P.491

MONE ZAIDI & SIMÓN MÉNDEZ-FERRER

The process of bone formation and maintenance requires the integrated action of several cell types. When this balance is perturbed, tumours can arise. In a paper published on page 491 of this issue, Yang *et al.*¹ assess the mutations underlying one such disorder, metachondromatosis, and in doing so identify a cell population that may be

^{*}This article and the paper under discussion¹ were published online on 17 July.

considered the first cancer stem cells to be discovered in bone.*

Long bones are initially formed from cells called chondrocytes, which create a flexible cartilage scaffold. Blood vessels then invade this mould, bringing cells that ultimately form bone marrow. Within the vessel walls are cells that give rise to osteoblasts, which mineralize the cartilaginous matrix to form a rigid structure — bone². Chondrocytes persist and proliferate at the tips of long bones to allow linear bone growth up

to puberty. Thereafter, bone is continuously remodelled, with packets of old bone being resorbed by osteoclasts and replaced by new bone formed by osteoblasts — a reparative process that maintains skeletal integrity throughout adult life³.

Uncontrolled chondrocyte proliferation in humans causes benign cartilage-tumour syndromes such as metachondromatosis. This heritable disease results mainly⁴ from loss-of-function mutations in *PTPN11*, a gene encoding the ubiquitously expressed phosphatase enzyme SHP2. But the cell population that becomes tumorigenic following the loss of SHP2 has hitherto remained a mystery. To investigate this, Yang *et al.* deleted *Ptpn11* in mice using two systems: one of these linked gene deletion to the expression of *LysM*, a gene expressed by osteoclasts, macrophages and monocytes (the latter two are white blood cells); the other linked deletion to the expression of *Ctsk*, which was thought to be an osteoclast-specific gene. The authors observed mildly reduced bone resorption, compared with normal animals, in the first category of mice — as would be expected following alteration to normal osteoclast activity. However, the *Ctsk*-dependent *Ptpn11* gene deletion resulted in mice with characteristics of human metachondromatosis, including the formation of new bone over existing bone.

This intriguing observation suggested that the *Ctsk*-mediated *Ptpn11* excision had in fact occurred not in osteoclasts but in another, previously uncharacterized, cell type. The authors identified these cells as belonging to the perichondrium — the layer of tissue that surrounds the cartilage at the tips of long bones. The perichondrial ring was known to be a source of chondrocyte precursors, but the identity and function of these progenitor cells was previously unclear. By documenting that *Ptpn11*-excised cells migrate from the perichondrium towards joint-associated cartilage, Yang and colleagues provide evidence that not only implicates these cells as chondrocyte progenitors but also indicates that they have a broad role in cartilage physiology.

An equally significant outcome of this study is the identification of a new site of action of *Ctsk* (cathepsin K), a protease enzyme that was thought to be expressed predominantly in osteoclasts and to function in bone-matrix degradation⁵. The finding that *Ctsk* is strongly expressed in perichondrial progenitor cells, and may therefore be involved in cartilage degradation, means that drugs inhibiting this protein (which are currently being developed to treat osteoporosis) could conceivably affect cartilage homeostasis. The results also mean that we must reconsider the use of *Ctsk* expression as a way of selectively modifying gene activity in osteoclasts⁶.

Interestingly, the *Ctsk*-expressing perichondrial cells express markers of mesenchymal stem cells, which are cells that can differentiate

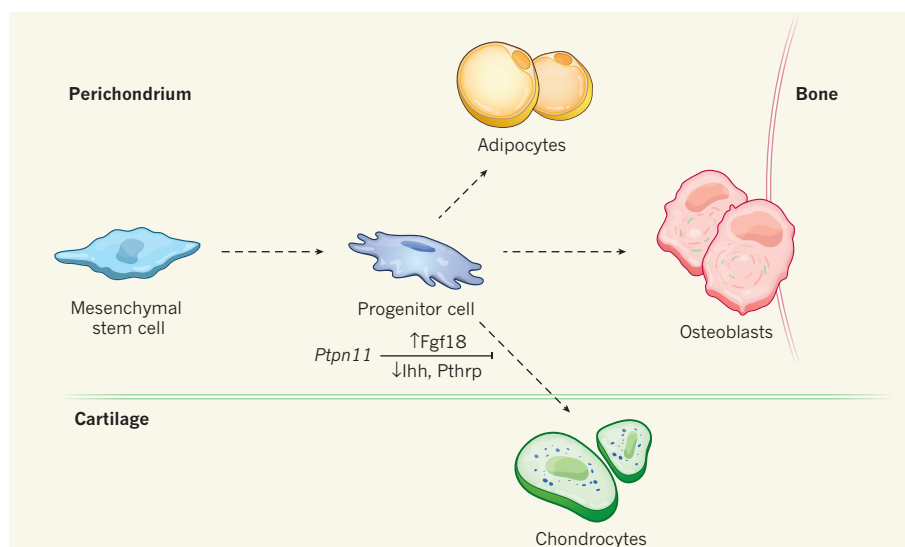


Figure 1 | Regulation of chondrocyte formation. Yang *et al.*¹ have identified, in mice, a population of progenitor cells in the perichondrium tissue layer that surrounds cartilage in long bones. These cells, which may derive from mesenchymal stem cells, can give rise to osteoblasts, adipocytes and chondrocytes *in vitro*. The authors show that the gene *Ptpn11* acts as a tumour suppressor in these cells, such that deletion of *Ptpn11* in these cells (not shown) initiates the formation of cartilage tumours similar to those seen in the human disorder metachondromatosis. The authors propose that this results from decreased production of the protein Fgf18 and, thereby, increased production of Ihh and Pthrp, which stimulate chondrocyte proliferation.

into several cell types, including osteoblasts, adipocytes and chondrocytes. The authors show that these perichondrial cells have similar multipotency *in vitro* (Fig. 1), and that the cells increase in number during disease progression in mice. These cells therefore seem to constitute a novel pool of Shp2-regulated, mesenchymal-lineage cells in mouse bone. However, it is not clear whether they fulfil the criteria for bona fide mesenchymal stem cells or simply represent more-committed skeletal precursor cells akin to others that function in bone remodelling^{7,8}. It is worth noting that both *Ctsk* and *Ptpn11* are highly expressed in some mesenchymal stem cells^{9,10}, suggesting a possible relationship between these cells and the newly identified perichondrial population.

Whether or not these perichondrial cells prove to be true mesenchymal stem cells, Yang and colleagues have for the first time established the presence of a ‘cancer stem cell’ in bone (even though the tumours that arise from these cells are benign). However, the method used in the present study will have led to *Ptpn11* deletion not only in the perichondrial progenitor cells, but also in cells that are derived from them, which raises the question of whether Shp2 deletion in the progenitors is the sole driver of tumour formation, or whether a ‘second hit’ of Shp loss is required. Addressing this question will be important for understanding the establishment and progression of the resulting disease.

In addition to causing metachondromatosis, germline mutations in human *PTPN11* cause Noonan, Noonan-like and LEOPARD syndromes, and somatic (non-germline)

mutations in this gene are associated with leukaemias and solid cancers. In metachondromatosis and LEOPARD syndrome, the mutations lead to inactivation of SHP2, whereas in the other cases SHP2 is abnormally activated. How can SHP2 signalling sometimes promote cell proliferation but in other cases, such as in the progenitor cells identified by Yang *et al.*, suppress it? It is known that *PTPN11* mutations can change the conformation of SHP2 in ways that either activate or inhibit oncogenic pathways. In LEOPARD syndrome, for example, *PTPN11* mutations result in a catalytically inactive SHP2 conformation that inhibits Erk activation^{11,12}, but although this would be expected to reduce oncogenesis, LEOPARD syndrome is associated with a high cancer incidence¹¹.

Similarly, Yang and colleagues found that Erk activation was inhibited in mice with *Ctsk*-driven *Ptpn11* deletion. They also showed that the cartilage tumorigenesis in these mice was the result of reduced secretion of the protein Fgf18, which led to elevated production of the proteins Ihh and Pthrp, both of which are potent stimulators of chondrocyte proliferation (Fig. 1). Treating the mice with a drug that targets the Ihh receptor effectively prevented tumour formation. Whether similar pathways drive tumour formation in human patients with metachondromatosis remains to be determined, but the collective results provide several targets for further investigation into the biology of this disease and its potential treatment. ■

Mone Zaidi and Simón Méndez-Ferrer
are in the Mount Sinai Bone Program, Icahn School of Medicine at Mount Sinai, New

York, New York 10029, USA. **S. M.-F.** is also at the Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain. e-mails: mone.zaidi@mountsinai.org; simon.mendez-ferrer@cnic.es

1. Yang, W. *et al.* *Nature* **499**, 491–495 (2013).
2. Maes, C. *et al.* *Dev. Cell* **19**, 329–344 (2010).
3. Zaidi, M. *Nature Med.* **13**, 791–801 (2007).
4. Bowen, M. E. *et al.* *PLoS Genet.* **7**, e1002050 (2011).

5. Zaidi, M., Troen, B., Moonga, B. S. & Abe, E. J. *Bone Miner. Res.* **16**, 1747–1749 (2001).
6. Nakamura, T. *et al.* *Cell* **130**, 811–823 (2007).
7. Park, D. *et al.* *Cell Stem Cell* **10**, 259–272 (2012).
8. Zaidi, M., Sun, L. & Blair, H. C. *Cell Stem Cell* **10**, 233–234 (2012).
9. Mendez-Ferrer, S. *et al.* *Nature* **466**, 829–834 (2010).
10. Isern, J. *et al.* *Cell Rep.* **3**, 1714–1724 (2013).
11. Grossmann, K. S., Rosário, M., Birchmeier, C. & Birchmeier, W. *Adv. Cancer Res.* **106**, 53–89 (2010).
12. Kontaridis, M. I., Swanson, K. D., David, F. S., Barford, D. & Neel, B. G. *J. Biol. Chem.* **281**, 6785–6792 (2006).

nor most of the mass, which is in the cold molecular phase.

It is only in the past few years that we have had the capability to detect emissions from the mass-laden molecular phase of outflows in nearby galaxies. This is because of increases in sensitivity made possible by updated technology on telescopes such as the Very Large Array in New Mexico and the IRAM Plateau de Bure interferometer in France, and with the advent of Europe's Herschel Space Observatory and now the Atacama Large Millimeter Array (ALMA) in Chile. One of the surprising results of these studies is that massive molecular winds are much more common than previously thought. However, many unanswered questions remain, meaning that it will be essential to make detailed investigations of the nearest examples of these winds.

In their paper (page 450), Bolatto and colleagues¹ report on observations, made with ALMA during its first year of operations, of one of the nearest starburst galaxies with a well-known superwind, NGC 253. This allowed them to image the cold carbon monoxide (CO) gas emission in NGC 253 at an incredible resolution of 50 parsecs. The unprecedented sensitivity of the researchers' observations revealed a number of very faint CO emission streamers out to heights of 120–320 parsecs for the first time. The morphology of these filaments closely matches that of the known ionized gas wind, providing clear evidence that the outflow also contains a substantial molecular component.

The authors also found that these prominent molecular-gas features seem to be connected

ASTROPHYSICS

How to catch a galactic wind

Observations obtained by the Atacama Large Millimeter Array in Chile's Atacama Desert have revealed properties of the cold molecular phase of the wind that is being blown out of a nearby starburst galaxy. **SEE LETTER P.450**

MARK WESTMOQUETTE

In this issue, Bolatto *et al.*¹ present ground-breaking interferometric observations of carbon monoxide emission from the nucleus and the central 2-kiloparsec region of a nearby starburst galaxy. The high sensitivity and resolution of these observations allowed the authors to detect faint molecular gas associated with a known superwind and to measure its mass-outflow rate for the first time.

When galaxies interact or merge, large quantities of gas can sometimes be channelled inwards as a result of the galaxies' gravitational forces. Under the right circumstances, this can trigger a massive burst of star formation in the heart of either or both galaxies, known as a starburst. The energy released, or fed back, by all these young stars in the form of radiation and winds (fast-moving gas particles blown off the stars) can have a marked effect on the remaining gas in the vicinity.

A single, short-lived star-formation event in a galaxy has an effect similar to an explosion. The energy released by the new stars creates a shock wave that compresses the surrounding gas into an expanding bubble. If the bubble inflates to a size comparable to that of the galaxy itself, then it is referred to as a superbubble. However, if the starburst proceeds for an extended period, then the initial bubble inflates and bursts, and the ongoing injection of energy drives the gas into a continuous flow of material called a galactic wind or superwind².

The exact mechanisms of how the stars accelerate and drive gas outwards are hotly debated topics^{3,4}. Constraints on theoretical predictions can come only from observations, and, because galactic winds are composed of

materials of mostly very low density that range over about five orders of magnitude in temperature, meaningful observations are difficult to make. Until recently, most of our understanding of galactic winds has come from the warm (10,000 kelvin) ionized and cooler (5,000 K) neutral gas phases of the winds, because they emit or absorb light in the easily accessible optical part of the spectrum. However, these phases contain neither most of the energy, which is in the hot X-ray-emitting phase,

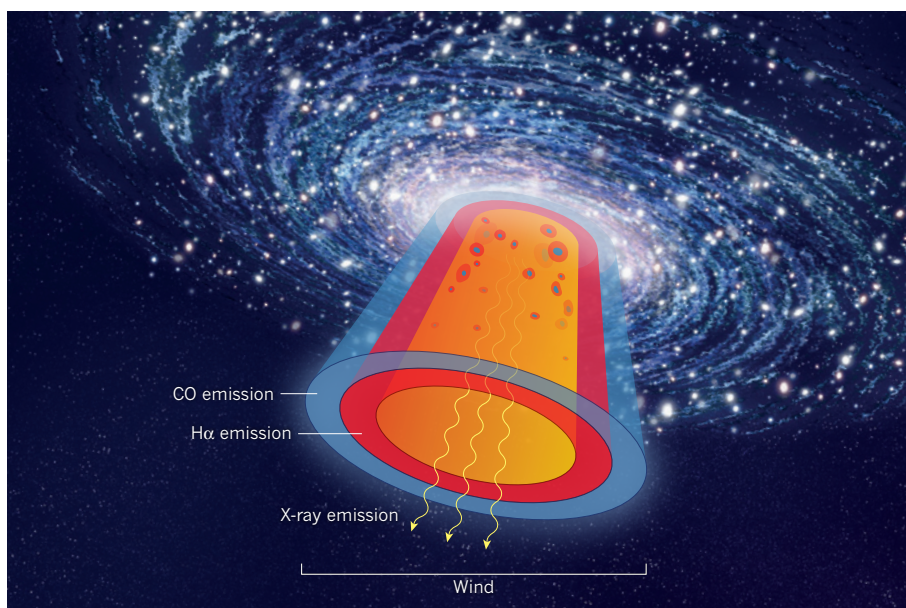


Figure 1 | The structure of the NGC 253 superwind. The wind being blown out of NGC 253 by the central starburst is thought to be approximately conical in structure with a temperature-stratified nature. The centre is filled with fast-moving and hot (a temperature of about 10^6 K) X-ray-emitting gas, and is surrounded by layers of cooler (10,000 K) $H\alpha$ -emitting gas and cold (100 K) molecular CO-emitting gas. Bolatto and colleagues' results¹ clearly show this cold CO layer at the edges of the warmer phases. The hot phase is also thought to contain clouds of cold material swept up by the wind⁶.

to the previously detected⁵ expanding molecular shells located on either side of the starburst region, thus providing an additional link between the starburst and the wind. The idea, built up from previous studies of winds, is that the hot wind fluid (temperatures greater than about 10^6 K), driven by the cumulative power of all the newly formed stars, interacts with the gas in the galaxy's disk and halo as it flows outwards. It is in the progressively cooler gas layers surrounding and embedded in the hot wind that the optical hydrogen- α (H α) emission (tracing the warm ionized gas at temperatures of about 10,000 K) and the molecular CO emission (tracing cold gas at temperatures of roughly 100 K) originates (Fig. 1).

The sensitivity of the authors' observations also allowed them to make direct measurements of the molecular-outflow speed and outflow rate. They found that the total mass-outflow rate is more than three times the measured star-formation rate. A considerable uncertainty exists in this measurement owing to the difficulties in converting from the measured outflowing CO mass to the predicted total outflowing mass, and in applying the correct geometric corrections (for example, for inclination), but it seems certain that the outflow rate is greater than the star-formation rate.

These results highlight several unanswered questions. The first relates to the depletion rate of gas and the starburst lifetime: if it is blowing out large amounts of gas, how long can the starburst continue given that it is removing the fuel for future star formation (that is, quenching itself)? Second, how is the energy required to eject this much material generated?

The mass-outflow rate sets constraints on the physical mechanisms needed to expel this material and thus forms important inputs for new theoretical models. Finally, how much of this ejected material is re-accreted later? One of the biggest unknowns with regard to starburst outflows is whether any of the ejected gas actually escapes the galaxy altogether. If it is not driven out far enough it may linger in the halo of the galaxy, cool, and rain back at a later time, potentially giving rise to a future starburst event, as might be happening in the starburst galaxy M82.

Although the molecular CO gas in the nuclear regions of NGC 253 has been imaged before⁵, the study by Bolatto *et al.* represents a major step forward in sensitivity and resolution, and provides new results on the wind's properties. It also showcases the revolutionary capabilities of ALMA in only its first year, operating with just 16 out of an eventual 66 antennas. ■

Mark Westmoquette is at the European Southern Observatory, 85748 Garching, Germany.

e-mail: westmoquette@gmail.com

1. Bolatto, A. D. *et al.* *Nature* **499**, 450–453 (2013).
2. Veilleux, S., Cecil, G. & Bland-Hawthorn, J. *Annu. Rev. Astron. Astrophys.* **43**, 769–826 (2005).
3. Murray, N., Ménard, B. & Thompson, T. A. *Astrophys. J.* **735**, 66 (2011).
4. Hopkins, P. F., Quataert, E. & Murray, N. *Mon. Not. R. Astron. Soc.* **421**, 3522–3537 (2012).
5. Sakamoto, K. *et al.* *Astrophys. J.* **636**, 685–697 (2006).
6. Strickland, D. K., Heckman, T. M., Weaver, K. A., Hoopes, C. G. & Dahlem, M. *Astrophys. J.* **568**, 689–716 (2002).

STRUCTURAL BIOLOGY

Meet the B family

The first crystal structures of class B G-protein-coupled receptors have been solved. They reveal features that might inform drug-development strategies for diseases ranging from osteoporosis to diabetes. SEE ARTICLES P.438 & P.444

PATRICK M. SEXTON & DENISE WOOTTEN

G-protein-coupled receptors (GPCRs) are the largest group of cell-surface receptors and are major targets for drug development^{1,2}. These proteins are characterized by a common architecture of seven transmembrane-spanning helical domains, and can be subdivided into three main groups: classes A, B and C. High-resolution structures of the membrane-spanning domain of GPCRs — the conduit for transmission of extracellular signals to the inside of a cell — provide snapshots that indicate how activating

and inactivating ligands modify the receptor structure. Until now, however, such studies have been principally restricted to class A receptors. In this issue, Hollenstein *et al.*³ (page 438) and Siu *et al.*⁴ (page 444) present the structures of the transmembrane domains of two class B members: corticotrophin-releasing factor-1 receptor and the glucagon receptor, respectively.*

Class B GPCRs include receptors for several peptide hormones, which are involved in a host of physiological functions from bone maintenance and glucose regulation to immune function and pain transmission. As a result, these receptors are targets for existing drugs that treat several disorders, including

*This article and the papers under discussion^{3,4} were published online on 17 July.

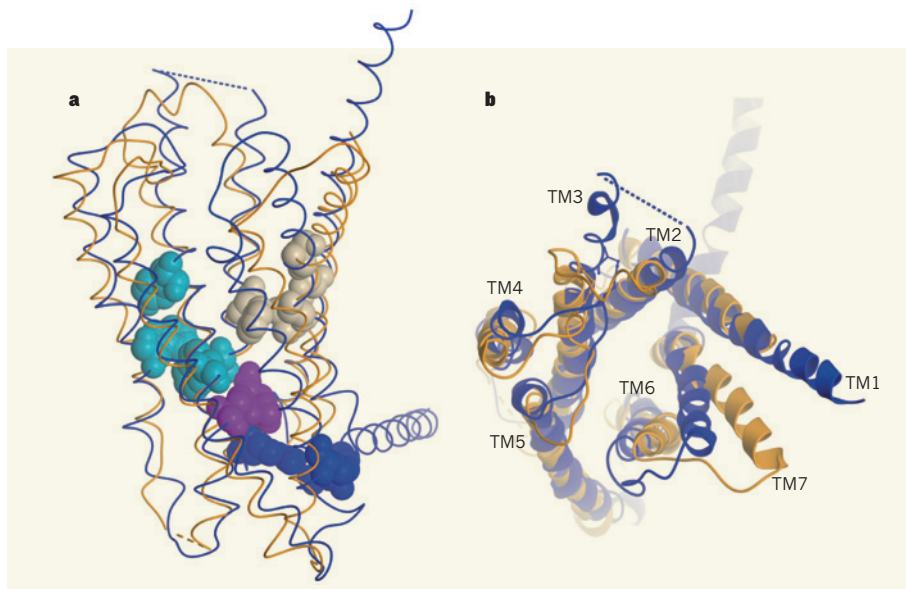


Figure 1 | Structural features of class B GPCRs. Hollenstein *et al.*³ and Siu *et al.*⁴ present the crystal structures of two class B G-protein-coupled receptors: CRF₁R (orange ribbons) and GCGR (blue ribbons), respectively. **a**, The structures reveal the locations of conserved amino-acid residues that form similar interactions in the two receptors, including between the transmembrane helices TM2, TM3 and TM4 (cyan), TM2 and TM3 (purple), TM1, TM2 and TM7 (beige), and TM2 and TM6 with the intracellular helix 8 (blue). **b**, The view of the proteins from outside the cell highlights the differences between the two structures at their extracellular faces, particularly in TM6 and TM7.

osteoporosis and type 2 diabetes, and are being actively pursued as targets for treating many more, from obesity and migraine to depression and chronic obstructive pulmonary disease.

Hollenstein and colleagues present a 3.0-ångström-resolution structure of the corticotrophin-releasing factor-1 receptor (CRF₁R) in complex with a small-molecule inhibitor. They arrived at this structure by introducing 12 thermostabilizing mutations into this GPCR and inserting the protein T4 lysozyme into its second intracellular loop. Siu and co-workers produced their 3.4-Å-resolution structure of the glucagon receptor (GCGR) using a version of the protein that was largely unmodified, except that its amino-terminal domain had been replaced with a thermally stabilized protein. The native N-terminal domain of class B GPCRs is crucial for peptide binding, but both teams removed this region to aid crystallization of the proteins.

As predicted, the core of both structures features seven transmembrane helices (TM1–TM7). However, although the relative positions of these helices at the intracellular face of the proteins overlap with those in class A GPCRs, there is substantial deviation between the two classes at the extracellular face. In both class B proteins, there are differences in the positioning of TM6 and TM7 that result in TM6 being shifted away from TM5, with TM1 seeming to move in parallel with TM7. This results in a wider and deeper extracellular cavity in the receptor core of the class B proteins that presumably forms part of the peptide-binding site. In addition, there are differences between the CRF₁R and GCGR structures themselves, in the upper segments of TM6 and

TM7 (Fig. 1). Although it is unclear whether these differences were influenced by the crystallization process, they indicate that the solution of transmembrane-core structures for other class B receptors will be required to help us understand how ligands bind and activate these proteins.

A major obstacle for the therapeutic targeting of class B receptors has been their notorious intractability for the identification of small-molecule ligands, in particular, small-molecule activators. The new structures shed light on why this is so: the openness of the receptors' binding pocket makes it difficult for a small ligand to engage sufficient key amino-acid residues to initiate activation of the receptor. Nonetheless, the solved structures show distinct subpockets that could represent sites for structure-based drug design.

Intriguingly, Hollenstein and colleagues' structure shows that the small-molecule inhibitor binds to a very deep pocket in the intracellular half of the CRF₁R core. This ligand forms extensive contacts with residues in TM3, TM5 and TM6, and presumably inhibits receptor activation by tethering the cytoplasmic half of TM6 to TM3 and TM5, thereby restricting conformational rearrangement of the intracellular face. This represents a new target for the design of small-molecule ligands. However, the amino-acid side chains in the equivalent region in the GCGR structure are more compact and would require reorganization to allow similarly sized ligands to bind.

The evolutionarily conserved amino-acid motifs in class A receptors have an important role in maintaining the receptors in an inactive (or weakly active) state. Although the

intracellular face of the class B receptors is similar to that seen for class A proteins (with the exception of an inward shift of TM7), some of the interactions that maintain the inactive class A conformations (including the ionic lock that tethers the cytoplasmic half of TM3 to TM6, the CWXP motif in TM6 and the NPXXY motif in TM7) are not present in the two class B receptors studied.

Class B receptors also have a distinct pattern of conserved amino-acid motifs that are important for maintenance of the inactive conformation and/or for conformational transitions required for activation. The CRF₁R and GCGR structures suggest conserved interactions between some of these key residues (Fig. 1). In addition, similar regions of contact are present between TM1 and TM2, TM1 and TM7, TM3 and TM4, and TM3 and TM6 in structures of both class A and B, although these interactions are mediated by different patterns of residues in each class. Thus, the new structures suggest that the two classes of proteins use distinct mechanisms for conformational control.

Although these reports represent a tremendous breakthrough in GPCR biology, as with all crystal structures, the intramembraneous class B structures provide only a snapshot of the receptors, which in reality are known (from cysteine-trapping studies⁵) to be highly dynamic proteins. Important questions remain about the final orientations of the N-terminal domains and transmembrane helices of the receptors, and about how natural activator molecules engage with both domains to activate the receptors. Answering these questions will require both crystallization of an intact ligand–receptor–G-protein complex and studies of receptor dynamics. ■

Patrick M. Sexton and Denise Wootten are at the Monash Institute of Pharmaceutical Sciences, Monash University, Parkville 3052, Victoria, Australia.

e-mails: patrick.sexton@monash.edu; denise.wootten@monash.edu

1. Fredriksson, R., Lagerström, M. C., Lundin, L. G. & Schiöth, H. B. *Mol. Pharmacol.* **63**, 1256–1272 (2003).
2. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. *Nature Rev. Drug Discov.* **5**, 993–996 (2006).
3. Hollenstein, K. *et al. Nature* **499**, 438–443 (2013).
4. Siu, F. Y. *et al. Nature* **499**, 444–449 (2013).
5. Dong, M. *et al. FASEB J.* **26**, 5092–5105 (2012).

CORRECTION

In the News & Views article 'Cloaking of heat' by Ulf Leonhardt (*Nature* **498**, 440–441; 2013), it was stated that "Einstein received his Nobel prize for his theory of diffusion, not for relativity." In fact, he received the prize "for his services to Theoretical Physics, and especially for his discovery of the law of the photoelectric effect".

Studies of the receptor-binding characteristics of H7N9 strains^{1–5} provide a plausible interpretation for these transmission findings. The authors used various methods to show that the H7N9 viruses, and A/Anhui in particular, bind efficiently to both $\alpha 2,3$ -linked and $\alpha 2,6$ -linked receptors. The report by Xiong *et al.*⁴ is particularly comprehensive, containing analyses of binding strength and specificity, and high-resolution X-ray crystallography structural data. These authors present the structures of A/Anhui HA, and a strictly avian H7 HA, in complexes with either human or avian receptor-like compounds. One notable feature they describe is that the human receptor binds the A/Anhui HA in an unusual orientation that has not been observed in previous structural studies of HA binding.

For several other HA subtypes, specific mutations that alter receptor specificity have been identified in the HA binding site, and structural studies have suggested mechanisms by which specific amino-acid residues in certain structural features of the protein — the 220-loop, 190-helix, 130-loop and 150-loop — might influence receptor specificity^{2,7–11} (Fig. 1). The A/Anhui HA contains a glutamine-to-leucine mutation at position 226 in the 220-loop, which is often associated with human adaptation, but it seems that additional changes, such as at position 228, will be required for these H7N9 viruses to switch completely to human-receptor specificity^{2,12}.

The dual binding specificity of these H7N9 strains poses the interesting question of why these viruses do not transmit efficiently in humans despite recognizing human receptors. One clue comes from comparing A/Anhui with an H5 virus that binds human receptors less strongly⁴ than A/Anhui HA, yet is efficiently transmitted among ferrets by aerosols¹³. The crucial difference might relate to the fact that this H5 strain shows virtually no binding to avian receptors⁴ and therefore might evade the inhibiting effects of extracellular mucin glycoproteins in the airways — these mucins are rich in $\alpha 2,3$ -linked glycans, which may inhibit avian viruses or viruses with dual specificity¹⁴ by sticking to them and effectively mopping them up before they infect cells.

Cumulatively, the new information suggests that H7 viruses may be within an amino-acid change or two of acquiring the capacity to preferentially recognize receptors in the upper airways of humans. However, the nuances of binding strength, dual recognition of avian and human receptors, and the requirement for balanced HA and NA functions make specific predictions difficult with regard to the imminent emergence of novel strains in humans. Furthermore, there are other viral genetic factors associated with human adaptation, and some of these are not well characterized. However, as these recent papers exemplify, evolving influenza viruses are continuing to test the waters of human adaptation. ■

David A. Steinhauer is in the Department of Microbiology and Immunology, Emory University School of Medicine, Atlanta, Georgia 30322, USA.
e-mail: dsteinh@emory.edu

1. Belser, J. A. *et al.* *Nature* <http://dx.doi.org/10.1038/nature12391> (2013).
2. Tharakaraman, K. *et al.* *Cell* **153**, 1486–1493 (2013).
3. Watanabe, T. *et al.* *Nature* <http://dx.doi.org/10.1038/nature12392> (2013).
4. Xiong, X. *et al.* *Nature* **499**, 496–499 (2013).
5. Zhou, J. *et al.* *Nature* **499**, 500–503 (2013).
6. Zhu, H. *et al.* *Science* **341**, 183–186 (2013).

7. Gamblin, S. J. *et al.* *Science* **303**, 1838–1842 (2004).
8. Ha, Y., Stevens, D. J., Skehel, J. J. & Wiley, D. C. *Proc. Natl Acad. Sci. USA* **98**, 11181–11186 (2001).
9. Lin, Y. P. *et al.* *Proc. Natl Acad. Sci. USA* **109**, 21474–21479 (2012).
10. Liu, J. *et al.* *Proc. Natl Acad. Sci. USA* **106**, 17175–17180 (2009).
11. Tharakaraman, K. *et al.* *Cell* **153**, 1475–1485 (2013).
12. Connor, R. J., Kawakita, Y., Webster, R. G. & Paulson, J. C. *Virology* **205**, 17–23 (1994).
13. Herfst, S. *et al.* *Science* **336**, 1534–1541 (2012).
14. Couceiro, J. N., Paulson, J. C. & Baum, L. G. *Virus Res.* **29**, 155–165 (1993).

CONDENSED-MATTER PHYSICS

Dark and heavy

Whenever a medium's dispersion and nonlinear properties appear hand in hand, particle-like entities known as solitons can form. These have now been observed in a gas of strongly interacting fermionic atoms. SEE ARTICLE P.426

CHRISTOPH BECKER

In 1834, the Scottish engineer John Scott Russell made a fascinating discovery. He observed a solitary water wave that travelled along a narrow canal for several miles without any significant change in its shape or amplitude. Ever since, such solitary waves, known as solitons, have been thought to have a key role in fundamental transport processes in myriad nonlinear systems — ranging from fibre-optic and meteorological systems to biological and astrophysical ones. On page 426 of this issue, Yefsah *et al.*¹ demonstrate the first experimental realization of 'dark solitons' in an ultracold quantum gas of strongly interacting fermionic atoms. Strikingly, they observe that the dynamics of a soliton in such a gas is slowed down by almost a factor of 20 compared with state-of-the-art theoretical predictions.*

Dark solitons are localized dips in the density distribution of a background medium. Like their bright analogues, they are particle-like entities because they retain their shape while they propagate or interact with other solitons². This behaviour is the result of a delicate balance of the medium's dispersion and nonlinear properties, and is in strong contrast to that of an ordinary wave packet (a superposition of plane waves) propagating in a linear medium, in which the packet will continuously spread and ultimately disappear. Although the appearance of solitons is quite a robust phenomenon and is not crucially dependent on the exact properties of the medium, the dynamics of solitons changes depending on the medium. In particular, for quantum-mechanical systems, thermal as well as quantum fluctuations perturb solitons

significantly, which makes solitons excellent probes of the system.

Dark solitons entered the world of ultracold quantum gases about a decade ago, when they were first observed³ in Bose–Einstein condensates (BECs). If a gas of bosons (particles with integer spin) is cooled to very low temperatures, the bosons tend to collectively condense into the lowest available quantum-mechanical state, giving rise to a BEC. One feature of a BEC is the emergence of frictionless flow, or superfluidity, which in turn can lead to the appearance of dark solitons. By exploiting the exceptional control attainable in BEC experiments, researchers have achieved several breakthroughs, including confirming the particle-like nature of solitons and the elasticity of soliton collisions^{4–6}.

The situation changes drastically for a gas of fermions — particles with half-integer spin. Unlike bosons, fermions are subject to the Pauli exclusion principle, according to which no two identical fermions may occupy the same quantum state simultaneously. To condense into a common quantum state with a macroscopic number of particles and form a superfluid, fermions first have to turn into bosons, and they can do so by forming pairs that have a resulting integer spin. The size of a pair crucially depends on the interaction between the particles, and determines the underlying physics. If the size is small compared with the inter-particle spacing, tightly bound molecules form and the system condenses into a BEC. If the size is much larger than the inter-particle distance, weakly bound pairs form and superfluidity ensues — much like the formation of Cooper pairs of electrons in superconductors. This second regime is known as the Bardeen–Cooper–Schrieffer (BCS) regime.

Theory can conveniently describe these two regimes, and the existence of solitons naturally

*This article and the paper under discussion¹ were published online on 17 July.

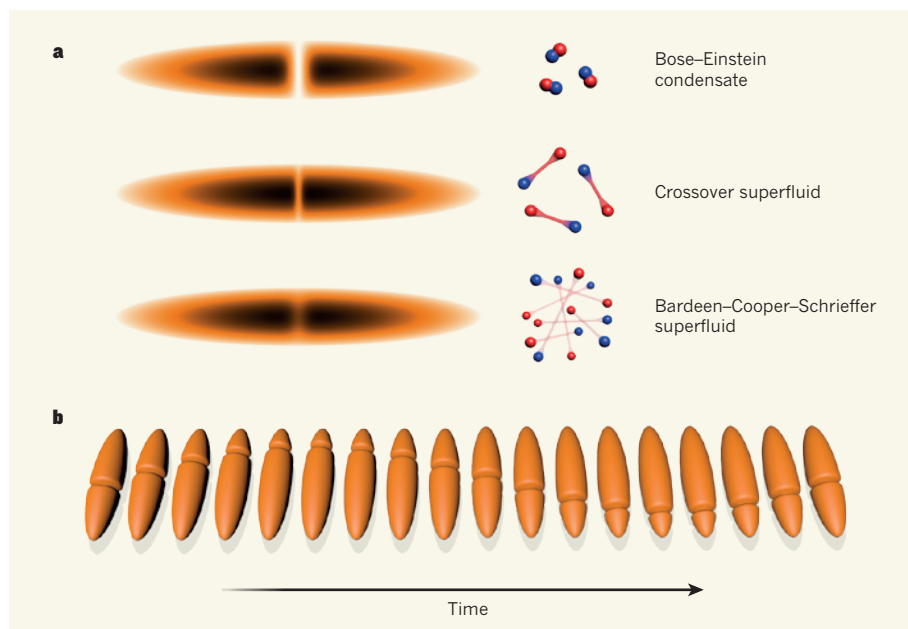


Figure 1 | Dark solitons in a fermionic superfluid. Yefsah *et al.*¹ created and observed the motion of dark solitons in an ultracold gas of strongly interacting fermionic atoms. **a**, The size of the atomic pairs (blue and red) comprising the gas determines how much such a soliton modulates the density of the gas (orange and black, with black denoting higher density and orange lower). The dark soliton (dip in the gas density) gets increasingly filled with non-condensed-gas atoms as the system moves from the Bose-Einstein condensation regime to the intermediate crossover superfluid, and then to a Bardeen-Cooper-Schrieffer superfluid. **b**, The propagation of a dark soliton in a trapped interacting Fermi gas.

arises from the existence of superfluidity in these interacting fermionic systems^{7–9}. In the crossover between the BEC and BCS regimes, the pair size is comparable to the inter-particle spacing and the gas is difficult to describe¹⁰. However, this regime is particularly appealing because its physics is the same for all different types of fermions — such a Fermi gas is said to be universal or unitary. Lacking any satisfactory theoretical description, researchers have conducted several experiments in this regime using ultracold quantum gases of fermions, and fundamental equilibrium and a few dynamical properties have been observed. One striking example was the observation of vortices, quantized units of angular momentum, which unambiguously confirmed the existence of robust superfluidity also in the unitary Fermi gas¹¹.

In their experiments, Yefsah *et al.* used the dynamical evolution of dark solitons across the BEC–BCS transition as a probe of the underlying Fermi gas. They continuously varied the pair size in an ultracold gas of lithium-6 fermionic atoms over the complete BEC–BCS crossover (Fig. 1) using a tool known as magnetic Feshbach resonance. They created dark solitons and allowed them to oscillate in the ‘harmonic trap’ they used to confine the atoms. Unexpectedly, they found that, whereas soliton oscillations deep in the BEC regime behaved quite as expected from theory, their motion was slowed down by almost a factor of 20 in the unitary and BCS regimes. The authors observed that, as these regimes are approached, the density dip that makes up a dark soliton gets

significantly more filled with non-condensed-gas atoms than expected from calculations. This filling makes the soliton heavier and thus slows down its motion^{7–9}.

Thermal and quantum fluctuations could both be responsible for these non-condensed atoms and thus for the existence of such heavy

solitons. Yefsah *et al.* have carefully analysed soliton oscillations at different temperatures and showed that the observed increase in the effective mass of the solitons is not caused by thermal fluctuations. On the other hand, existing theories underestimate the effect of quantum fluctuations, so Yefsah and colleagues’ study should be regarded as a benchmark quantum simulation to test future theories of strongly interacting Fermi gases. Regardless of whether exotic states inside the solitons, such as Andreev bound states, or completely different and as yet unknown mechanisms can generate the slow motion of the dark solitons, shedding light on this problem might turn out to be beneficial for our general understanding of strongly interacting fermions. ■

Christoph Becker is at the Institute for Laser Physics, University of Hamburg, 22761 Hamburg, Germany.
e-mail: cbecker@physnet.uni-hamburg.de

1. Yefsah, T. *et al.* *Nature* **499**, 426–430 (2013).
2. Zabusky, N. J. & Kruskal, M. D. *Phys. Rev. Lett.* **15**, 240–243 (1965).
3. Burger, S. *et al.* *Phys. Rev. Lett.* **83**, 5198–5201 (1999).
4. Becker, C. *et al.* *Nature Phys.* **4**, 496–501 (2008).
5. Weller, A. *et al.* *Phys. Rev. Lett.* **101**, 130401 (2008).
6. Stellmer, S. *et al.* *Phys. Rev. Lett.* **101**, 120406 (2008).
7. Busch, T. & Anglin, J. R. *Phys. Rev. Lett.* **84**, 2298–2301 (2000).
8. Scott, R. G., Dalfovo, F., Pitaevskii, L. P. & Stringari, S. *Phys. Rev. Lett.* **106**, 185301 (2011).
9. Spuntarelli, A., Carr, L. D., Pieri, P. & Stringari, G. C. *New J. Phys.* **13**, 035010 (2011).
10. Giorgini, S., Pitaevskii, L. P. & Stringari, S. *Rev. Mod. Phys.* **80**, 1215–1274 (2008).
11. Zwierlein, M. W., Abo-Shaeer, J. R., Schirotzek, A., Schunk, C. H. & Ketterle, W. *Nature* **435**, 1047–1051 (2005).

CELL BIOLOGY

Tumour stem cells in bone

Activation of the signalling molecule SHP2 is implicated in driving several cancers. In a newly described class of bone-progenitor cells, however, it seems that the protein acts as a tumour suppressor. SEE LETTER P.491

MONE ZAIDI & SIMÓN MÉNDEZ-FERRER

The process of bone formation and maintenance requires the integrated action of several cell types. When this balance is perturbed, tumours can arise. In a paper published on page 491 of this issue, Yang *et al.*¹ assess the mutations underlying one such disorder, metachondromatosis, and in doing so identify a cell population that may be

^{*}This article and the paper under discussion¹ were published online on 17 July.

considered the first cancer stem cells to be discovered in bone.*

Long bones are initially formed from cells called chondrocytes, which create a flexible cartilage scaffold. Blood vessels then invade this mould, bringing cells that ultimately form bone marrow. Within the vessel walls are cells that give rise to osteoblasts, which mineralize the cartilaginous matrix to form a rigid structure — bone². Chondrocytes persist and proliferate at the tips of long bones to allow linear bone growth up

INFLUENZA

Pathways to human adaptation

An outbreak of avian H7N9 influenza in humans was reported in early 2013. Structural and infection studies are helping to reveal how these viruses can adapt to infect, and potentially transmit in, new species. [SEE LETTERS P.496 & P.500](#)

DAVID A. STEINHAUER

The seasonal varieties of influenza that circulate through human populations are serious enough, but what we really fear is the emergence in humans of novel pandemic strains. Such viruses could emerge from natural avian hosts such as ducks, either directly or through intermediate hosts such as chickens or pigs. On the unpredictable occasions when these viruses cross species barriers and infect humans, the consequences can be dramatic: over the past decade, highly pathogenic H5N1 avian strains have caused limited but lethal human outbreaks, and the emergence of human H7N9 viruses in China earlier this year reinforced this threat. Several papers published over the past two months^{1–6}, including two in this issue (Xiong *et al.*, page 496 and Zhou *et al.*, page 500) have addressed receptor binding, transmission and pathogenesis of influenza viruses, with a focus on H7N9 strains.* These reports document the viral characteristics and associated genetic changes that partially explain why H7N9 strains can sporadically infect humans, and offer insights into the evolutionary pathways by which such strains might adapt

for efficient human-to-human transmission.

A vast genetic reservoir of influenza A strains exists in their natural avian hosts. The viruses are classified on the basis of the antigenic properties of their surface glycoproteins haemagglutinin (HA) and neuraminidase (NA), from which is derived the common nomenclature for influenza A subtypes — H5N1, H7N9 and so on. Of the 16 HA subtypes and 9 NA subtypes known to circulate in waterfowl, only H1N1, H2N2 and H3N2 have emerged and developed the capacity for efficient transmission in humans over the past century. Although adaptation for human transmission probably involves multiple genetic factors, one crucial requirement relates to the receptor-binding properties of the HA glycoprotein, which must mutate before it can recognize receptors in the human respiratory tract.

Influenza viruses initiate infection by HA-mediated binding to cell-surface molecules, such as glycoproteins and glycolipids, that contain carbohydrate chains terminating in sialic acid. As a general rule, avian viruses tend to bind to receptors with sialic acid linked to these chains by $\alpha,2,3$ glycosidic bonds, whereas human strains show a preference for receptors with $\alpha,2,6$ -linked sialic acid (Fig. 1). These preferences generally reflect the distribution of receptors at sites of infection in the virus's

preferred hosts and the natural routes of transmission. Human influenza viruses are transmitted by the respiratory route, and the human upper airway is rich in $\alpha,2,6$ -linked receptors. Avian viruses, by contrast, often use a faecal/oral transmission route, and the intestinal tract of ducks, for example, contains a high density of $\alpha,2,3$ -linked receptors. Therefore, a key issue in regard to host adaptation seems to be the mechanisms by which the structure of HA's receptor-binding sites can mutate to discriminate between 'avian' and 'human' receptors.

The recent studies on H7N9 influenza viruses used two human isolates, referred to as A/Anhui and A/Shanghai. Three of the studies^{1,3,6} investigated the infection, pathogenesis and transmission of H7N9 in experimental animals, including ferrets. In ferrets, human-adapted pandemic and seasonal influenza viruses are generally efficiently transmitted by the respiratory route in aerosols, whereas avian viruses are not; so ferrets are considered the best model for predicting a virus's capacity for human-to-human transmission. Notably, each of the three studies showed that approximately one third of the ferrets that had not previously been exposed to the H7N9 viruses became infected through respiratory droplets from infected ferrets, indicating that these viruses have an intermediate capacity to be transmitted by aerosols.

*This article was published online on 17 July and the papers under discussion were published online on 20 June⁴, 3 July⁵ and 10 July^{1,3}.

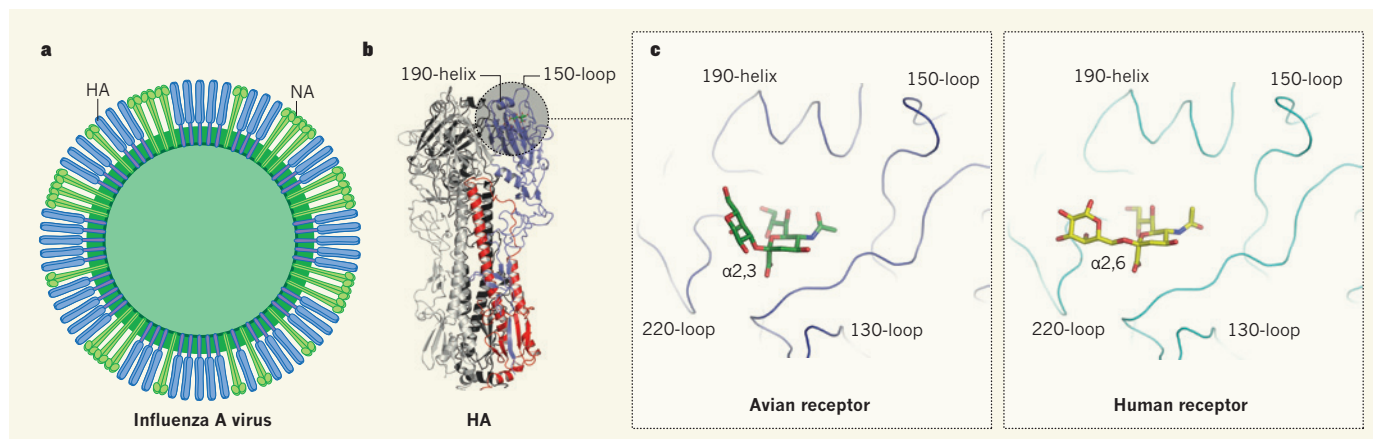


Figure 1 | Viral species preferences. **a**, In the influenza A virus, a layer of haemagglutinin (HA) and neuraminidase (NA) glycoproteins protrude from the viral membrane. **b**, HA is a trimeric molecule, and the membrane-distal domain of each monomer contains a binding site (shaded region) for the carbohydrate chains of glycolipids or glycoproteins with sialic acid at their ends. **c**, Influenza

viruses that circulate in birds tend to bind to receptors in which the sialic acid contains $\alpha,2,3$ glycosidic bonds, whereas human strains show a preference for receptors with $\alpha,2,6$ -linked sialic acid. Differences in the amino acids making up the structural features of the HA binding site (the 190-helix, 220-loop, 130-loop and 150-loop) are thought to influence this receptor specificity.

Studies of the receptor-binding characteristics of H7N9 strains^{1–5} provide a plausible interpretation for these transmission findings. The authors used various methods to show that the H7N9 viruses, and A/Anhui in particular, bind efficiently to both $\alpha 2,3$ -linked and $\alpha 2,6$ -linked receptors. The report by Xiong *et al.*⁴ is particularly comprehensive, containing analyses of binding strength and specificity, and high-resolution X-ray crystallography structural data. These authors present the structures of A/Anhui HA, and a strictly avian H7 HA, in complexes with either human or avian receptor-like compounds. One notable feature they describe is that the human receptor binds the A/Anhui HA in an unusual orientation that has not been observed in previous structural studies of HA binding.

For several other HA subtypes, specific mutations that alter receptor specificity have been identified in the HA binding site, and structural studies have suggested mechanisms by which specific amino-acid residues in certain structural features of the protein — the 220-loop, 190-helix, 130-loop and 150-loop — might influence receptor specificity^{2,7–11} (Fig. 1). The A/Anhui HA contains a glutamine-to-leucine mutation at position 226 in the 220-loop, which is often associated with human adaptation, but it seems that additional changes, such as at position 228, will be required for these H7N9 viruses to switch completely to human-receptor specificity^{2,12}.

The dual binding specificity of these H7N9 strains poses the interesting question of why these viruses do not transmit efficiently in humans despite recognizing human receptors. One clue comes from comparing A/Anhui with an H5 virus that binds human receptors less strongly⁴ than A/Anhui HA, yet is efficiently transmitted among ferrets by aerosols¹³. The crucial difference might relate to the fact that this H5 strain shows virtually no binding to avian receptors⁴ and therefore might evade the inhibiting effects of extracellular mucin glycoproteins in the airways — these mucins are rich in $\alpha 2,3$ -linked glycans, which may inhibit avian viruses or viruses with dual specificity¹⁴ by sticking to them and effectively mopping them up before they infect cells.

Cumulatively, the new information suggests that H7 viruses may be within an amino-acid change or two of acquiring the capacity to preferentially recognize receptors in the upper airways of humans. However, the nuances of binding strength, dual recognition of avian and human receptors, and the requirement for balanced HA and NA functions make specific predictions difficult with regard to the imminent emergence of novel strains in humans. Furthermore, there are other viral genetic factors associated with human adaptation, and some of these are not well characterized. However, as these recent papers exemplify, evolving influenza viruses are continuing to test the waters of human adaptation. ■

David A. Steinhauer is in the Department of Microbiology and Immunology, Emory University School of Medicine, Atlanta, Georgia 30322, USA.
e-mail: dsteinh@emory.edu

1. Belser, J. A. *et al.* *Nature* <http://dx.doi.org/10.1038/nature12391> (2013).
2. Tharakaraman, K. *et al.* *Cell* **153**, 1486–1493 (2013).
3. Watanabe, T. *et al.* *Nature* <http://dx.doi.org/10.1038/nature12392> (2013).
4. Xiong, X. *et al.* *Nature* **499**, 496–499 (2013).
5. Zhou, J. *et al.* *Nature* **499**, 500–503 (2013).
6. Zhu, H. *et al.* *Science* **341**, 183–186 (2013).

7. Gamblin, S. J. *et al.* *Science* **303**, 1838–1842 (2004).
8. Ha, Y., Stevens, D. J., Skehel, J. J. & Wiley, D. C. *Proc. Natl Acad. Sci. USA* **98**, 11181–11186 (2001).
9. Lin, Y. P. *et al.* *Proc. Natl Acad. Sci. USA* **109**, 21474–21479 (2012).
10. Liu, J. *et al.* *Proc. Natl Acad. Sci. USA* **106**, 17175–17180 (2009).
11. Tharakaraman, K. *et al.* *Cell* **153**, 1475–1485 (2013).
12. Connor, R. J., Kawakita, Y., Webster, R. G. & Paulson, J. C. *Virology* **205**, 17–23 (1994).
13. Herfst, S. *et al.* *Science* **336**, 1534–1541 (2012).
14. Couceiro, J. N., Paulson, J. C. & Baum, L. G. *Virus Res.* **29**, 155–165 (1993).

CONDENSED-MATTER PHYSICS

Dark and heavy

Whenever a medium's dispersion and nonlinear properties appear hand in hand, particle-like entities known as solitons can form. These have now been observed in a gas of strongly interacting fermionic atoms. SEE ARTICLE P.426

CHRISTOPH BECKER

In 1834, the Scottish engineer John Scott Russell made a fascinating discovery. He observed a solitary water wave that travelled along a narrow canal for several miles without any significant change in its shape or amplitude. Ever since, such solitary waves, known as solitons, have been thought to have a key role in fundamental transport processes in myriad nonlinear systems — ranging from fibre-optic and meteorological systems to biological and astrophysical ones. On page 426 of this issue, Yefsah *et al.*¹ demonstrate the first experimental realization of 'dark solitons' in an ultracold quantum gas of strongly interacting fermionic atoms. Strikingly, they observe that the dynamics of a soliton in such a gas is slowed down by almost a factor of 20 compared with state-of-the-art theoretical predictions.*

Dark solitons are localized dips in the density distribution of a background medium. Like their bright analogues, they are particle-like entities because they retain their shape while they propagate or interact with other solitons². This behaviour is the result of a delicate balance of the medium's dispersion and nonlinear properties, and is in strong contrast to that of an ordinary wave packet (a superposition of plane waves) propagating in a linear medium, in which the packet will continuously spread and ultimately disappear. Although the appearance of solitons is quite a robust phenomenon and is not crucially dependent on the exact properties of the medium, the dynamics of solitons changes depending on the medium. In particular, for quantum-mechanical systems, thermal as well as quantum fluctuations perturb solitons

significantly, which makes solitons excellent probes of the system.

Dark solitons entered the world of ultracold quantum gases about a decade ago, when they were first observed³ in Bose–Einstein condensates (BECs). If a gas of bosons (particles with integer spin) is cooled to very low temperatures, the bosons tend to collectively condense into the lowest available quantum-mechanical state, giving rise to a BEC. One feature of a BEC is the emergence of frictionless flow, or superfluidity, which in turn can lead to the appearance of dark solitons. By exploiting the exceptional control attainable in BEC experiments, researchers have achieved several breakthroughs, including confirming the particle-like nature of solitons and the elasticity of soliton collisions^{4–6}.

The situation changes drastically for a gas of fermions — particles with half-integer spin. Unlike bosons, fermions are subject to the Pauli exclusion principle, according to which no two identical fermions may occupy the same quantum state simultaneously. To condense into a common quantum state with a macroscopic number of particles and form a superfluid, fermions first have to turn into bosons, and they can do so by forming pairs that have a resulting integer spin. The size of a pair crucially depends on the interaction between the particles, and determines the underlying physics. If the size is small compared with the inter-particle spacing, tightly bound molecules form and the system condenses into a BEC. If the size is much larger than the inter-particle distance, weakly bound pairs form and superfluidity ensues — much like the formation of Cooper pairs of electrons in superconductors. This second regime is known as the Bardeen–Cooper–Schrieffer (BCS) regime.

Theory can conveniently describe these two regimes, and the existence of solitons naturally

*This article and the paper under discussion¹ were published online on 17 July.

CAREERS

EUROPE EU aims to ease migration for researchers and students **p.511**

PUBLISHING Flaws in papers are being caught more rapidly **p.511**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

OCEAN/CORBIS



BIOTECHNOLOGY

Independent streak

Scientists willing to take a risk are setting up individual research operations in rented lab space.

BY VIRGINIA GEWIN

Erhan Perlstein was frustrated. As a non-traditional postdoc, he had spent five years running an independent lab at Princeton University in New Jersey. He wanted to continue doing what he had trained for, but a tough academic job market meant that he had no guarantees. So he decided to move to a noted biotechnology hub, California's San Francisco Bay area, to try launching his own lab without the support of an academic institution.

His goal is to validate a drug-discovery protocol and start networking with potential investors. Perlstein explored several lab spaces, and in July he signed a two-month lease for a bench at the Molecular Sciences Institute, a non-profit research facility in Berkeley, California.

Perlstein is one of a growing number of young scientists who, stymied by the lack of jobs or the slow pace of research in academia, are pursuing cutting-edge research in their own spaces. Research tools and lab areas are becoming more affordable, particularly in the communal set-ups dubbed biohacker spaces, which seek to enable biological experimentation for hobbyists as well as for entrepreneurs. Funding schemes aimed at early-stage companies are also starting to offer a way to pursue an independent science career. Despite the risks, the option of launching an individual research operation is gaining traction throughout the United States and Europe.

Such routes, however, involve striking out into the unknown. Some non-profit, volunteer-driven hacker spaces are struggling to stay afloat, and it is not yet clear how independent research operations will become sustainable, especially if their work has no immediate practical applications. Young scientists eager for this brash brand of independence should consider their research needs, find out which spaces are optimally set up to help them, and determine how best to secure financial support.

HACKER CRAZE

The first biohacker spaces were fuelled by the do-it-yourself biology movement, in which untrained hobbyists started running biotechnology experiments in their kitchens and garages (see *Nature* **467**, 650–652; 2010).

These days, both amateurs and professionals are seeking out cheap lab space. The set-ups vary: hackers can rent a stretch of lab bench at GenSpace in New York, for example, for US\$100 per month including access to reagents and ►

► some biotechnology instruction. But scientists working on specific projects often have more specialized needs, and a new model is emerging in the United States. The hybrid ‘hackubator’ fuses the independence and affordability of hacker spaces with the entrepreneurial bent of biotech business incubators.

Bio, Tech and Beyond, a hackubator space that opened this month in Carlsbad, California, charges \$400 per month for a lab bench — much less than the roughly \$900 per month of conventional biotech incubators. Members get access to centrifuges, cell-culture incubators, plate robotics and three-dimensional printing, as well as help with writing grant applications.

“Biotech incubators offer company spaces, hackubators offer pre-company spaces — when it’s just you and an idea, but you need the equipment and connections to biotech and pharma to get an idea off the ground,” says Ryan Bethencourt, an entrepreneur involved in planning two biohacker spaces in the San Francisco area.

Several cities are encouraging the development of hacker spaces to create jobs and generate tax revenue. Joseph Jackson, co-founder of Bio, Tech and Beyond, got his organization’s space essentially rent-free from the city of Carlsbad; in exchange, he is supposed to help launch eight companies within two years. He already has eight pilot users and 12 further requests for space. “If we can’t make this model work here, it won’t work anywhere,” he says.

Brightwork CoResearch, a biosafety-level-2 facility for independent scientists, is set to open in August next to Rice University in Houston, Texas, home to one of the country’s largest medical centres and a fledgling biotech community. Brightwork’s initial funding comes from a stem-cell biologist and a local entrepreneur. There will be 20 benches dedicated to full-time researchers and another 20 for part-time researchers; almost one-quarter of the spaces have already been assigned. Most of the researchers hope to form companies, but sustaining their work won’t be easy. “The big question is, how will groups fund their research,” says co-founder Jacob Schiach. “The only thing that matters to me is that people are able to push the science they love forward.”

Europe’s budding biohacker scene includes La Paillasse in Paris: the continent’s largest biohacker space, with more than 30 members, who pay what they can for the space (even as little as €2, or US\$2.60) and use mostly donated

equipment. Co-founder and president Thomas Landrain, a PhD student at the Institute of Systems and Synthetic Biology in Evry, France, says there are ten projects currently in development; at least three, including one focused on producing cheap ink from microbes, could become start-up projects. La Paillasse is growing rapidly and, with public funding and support from the mayor of Paris, will soon move to a larger building in the city centre, says Landrain.

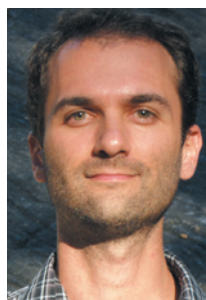
INDEPENDENTLY WEALTHY

Once they have found a space for their work, scientists still need to find funding streams. That may mean taking on contract research, using crowd-funding to get enterprises off the ground or winning speciality grants, such as the small business innovation research grants from the US National Institutes of Health and the US National Science Foundation.

Saul Griffith is the co-founder of OtherLab, an independent research lab in San Francisco. His company gets roughly 40% of its revenue from grants to conduct early-stage, high-risk applied research on projects ranging from solar energy to robotics for agencies such as the US Department of Energy and the Defense Advanced Research Projects Agency. Another 20% is from contract research for companies such as Ford and General Electric. The rest comes from developing intellectual property and selling products — notably the OtherMill, a small, computer-controlled mill that can cut everything from circuit boards to jewellery.

UBiome — launched from the California Institute for Quantitative Biosciences, a start-up incubator at the University of California, San Francisco — gets its funding in small chunks by selling microbiome-profiling kits. For \$89, it will sequence the genomes of the microbes living in a person’s mouth, gut, nose or genitals. It gives individuals insight into their bacteria, but co-founder Zachary Apte also wants to mine the data to uncover insights such as how eating habits influence microbiomes. “I was ready to do my own research and had my own ideas to explore, but it takes a long time to get there taking the academic path,” he says. His is not the only outfit funded by the public: earlier this year, three biohackers launched a crowd-funding campaign to raise the money to bioengineer a glowing plant (see *Nature* 498, 15–16; 2013). They made just under \$500,000.

A few funding schemes specifically provide early-stage support to high-risk, high-reward independent ventures. Schiach has just launched SynBio axlr8r, a programme to jumpstart synthetic-biology companies with funding from SOS Ventures, a venture-capital firm in Kinsale, Ireland. “The goal of SynBio axlr8r is to take an idea through proof of concept to form a company within 90 days,” he says. The first batch of 10–20 successful applicants will get US\$30,000 each and a space at a science incubator in Cork, Ireland.



“I don’t want to give up doing science and I’m willing to go the distance to find a way to make it work.”

Ethan Perlstein



Saul Griffith funds his independent research in areas such as robotics using grants and contract work.

Through a scheme called Breakout Labs, the Thiel Foundation in San Francisco funds early-stage companies working on radical, big-risk ideas that are unattractive to federal funding and venture capital. Grants of \$350,000 help recipients to reach technical milestones, at which point they can start to attract more conventional grants or investments. “I started a commercial enterprise because I wanted tools that didn’t exist, to conduct my research,” says Todd Huffman, who used a Breakout Labs grant to support 3Scan, an early-stage company that he co-founded to develop a knife-edge scanning microscope that simultaneously cuts and scans tissue slices to create three-dimensional models of samples. “I wanted to build these tools during my PhD and my adviser didn’t think it was a good use of time,” he says. Huffman dropped out of his PhD programme but maintains an interest in neuron morphology; after establishing 3Scan, he hopes eventually to return to his PhD with a new set of tools in hand.

Darren Zhu received a \$100,000 two-year grant from the Thiel Foundation to leave university and become an entrepreneur. He is exploring ways to engineer organisms to produce novel molecules, a project that could help him to find a niche as the pharmaceutical industry continues to outsource early-stage research and development. “At this stage of my life, I can afford to take risks, to swing for the fences and see where it takes me,” he says. In addition to his Thiel grant, he has money from the Bill & Melinda Gates Foundation in Seattle, Washington. He used it to buy lab equipment from liquidated firms, with which he has furnished an independent, industrial lab space in Mountain View, California.

OVERCOMING THE ODDS

Zhu says that one of the bigger hurdles of pursuing an independent path is that much of the biotechnology field still places a premium on “old-fashioned credentialing” — publishing work in prestigious journals or

having done a postdoc at a reputable institution. It is a challenge, he says, but one he has overcome by securing high-profile grants and fellowships.

The first challenge that Perlstein encountered was finding a space that met his needs. He was intrigued by the mix of hobbyists and professionals in biohacker spaces but, in the end, they just did not work for his specialized research. The space he eventually settled on is comparatively pricey at roughly \$2,000 per month — but it offers a fully stocked yeast lab that will allow him to hit the ground running, conducting the six weeks of experiments that he anticipates needing to develop a rapid yeast screen for rare-disease therapeutics.

Bethencourt is helping to plan the Counter Culture Labs hacker space in Oakland, California, which will provide lab space to at least 50 paying members from this autumn onwards. But, seeing the need for more specialized spaces, he has also co-founded a small therapeutics-focused hacking space, called Berkeley Biolabs, which is set to open this summer.

Perlstein notes further challenges such as finding insurance to cover mishaps at a rented lab bench. His biggest hurdle, however, may be determining whether his research will translate into a viable company — and whether he wants it to.

With funding and positions growing more difficult to find in academia, many more scientists may become willing to take on the risks. And biohackers looking to continue their work in the face of a dismal funding scene will ultimately form their own experimental units. “People are taking control,” says Perlstein. “I don’t want to give up doing science and I’m willing to go the distance to find a way to make it work.” ■

Virginia Gewin is a freelance writer based in Portland, Oregon.

EUROPE

Boost for mobility

The European Commission (EC) has proposed a strategy to boost movement into and out of the European Union (EU) for students and researchers. In an initiative outlined on 11 July, some visa restrictions would be eased to help non-European scientists to enter the EU and stay for more than 90 days. Member states will respond to the directive by August. The strategy also calls for a January relaunch of the Erasmus student-exchange scheme, with a budget of €14.5 billion (US\$19 billion) for 2014–20 — a 40% increase on 2007–13. The Marie Skłodowska-Curie Actions programme to fund international researchers will have €5.6 billion for 2014–20, but expects to recruit an extra 10,000 scientists with co-funding from participating countries, says EC policy officer Ragnhild Solvi Berg.

UNITED STATES

Second postdoc survey

In a follow-up to an influential study, the US National Postdoctoral Association (NPA) in Washington DC has launched a survey to gather data on postdoc pay and benefits, policies, services and demographics. Results will be out next year. In 2003, Sigma Xi in Research Triangle Park, North Carolina, along with the NPA and partners, ran the first US multi-campus survey on postdocs’ work, goals and perceptions of policies and practices at their institutions. The results informed institutional and federal policy. Lorraine Tracey, chair of the NPA board, says that the current survey will provide longitudinal data. “This will illustrate for federal legislators and funding agencies how policies can effect change across the postdoctorate,” she notes.

PUBLISHING

Retractions speed up

The number of retractions has risen in recent years in part because journals are acting more quickly, says a study in *PLOS ONE* (R. G. Steen *et al.* *PLOS ONE* 8, e68397; 2013). R. Grant Steen of MediCC! medical-communication consultants in Chapel Hill, North Carolina, and his colleagues analysed 2,047 papers published between 1973 and 2012, and later retracted. Those published in or before 2002 were retracted in 50 months on average; later, retractions took 24 months. Data that seemed “too good to be true” may now slip by less often, says Steen. “There might be a tendency to confront the author.”